Advanced Occupancy Measurement Using Sensor Fusion

Tobore Ekwevugbe

A thesis submitted in partial fulfilment of the requirements of De Montfort University for the degree of Doctor of Philosophy

Institute of Energy and Sustainable Development

De Montfort University, Leicester, United Kingdom

December, 2013.

DECLARATION

No part of the material described in this thesis has been submitted for the award of any other degree or qualification in this or any other university or college of advanced education.

Tobore Ekwevugbe

ACKNOWLEDGEMENT

I would like to express my deepest gratitude to my first supervisor, Dr Neil Brown, for his inspiring guidance and invaluable support throughout my work. I am also deeply grateful to other member of the supervisory team, Dr Vijay Pakka, Dr Simon Rees, and Dr Denis Fan whose comments and suggestions have greatly contributed to the success of this work.

I am thankful to many colleagues in the IESD for the excellent discussions, and all the good laughs, for continuous encouragement, and generally a great time, and with whom I have established friendships that I cherish. Special thanks go to Muhammad Mazhar, Dr Ivan Korolija, Dr Divine Novieto and Dr Kate Irvine.

I am hugely grateful to my parents, my brother and sisters for their endless, love, support and complete understanding all through the way. I also thank many friends that weave into my everyday life, special thanks also goes to Joanna Wilczynska, Franklin Ogbogu, Oluseyi Gbemikaiye, Mrs Belema Olisa, Dr Sorbarikor Lebura and Mrs Oluwatobiloba Akiyesi.

Without the financial support from De Monfort University, this work and thesis would never have been realised, and I am grateful for this support.

ABSTRACT

With roughly about half of the energy used in buildings attributed to Heating, Ventilation, and Air conditioning (HVAC) systems, there is clearly great potential for energy saving through improved building operations. Accurate knowledge of localised and real-time occupancy numbers can have compelling control applications for HVAC systems. However, existing technologies applied for building occupancy measurements are limited, such that a precise and reliable occupant count is difficult to obtain. For example, passive infrared (PIR) sensors commonly used for occupancy sensing in lighting control applications cannot differentiate between occupants grouped together, video sensing is often limited by privacy concerns, atmospheric gas sensors (such as CO_2 sensors) may be affected by the presence of electromagnetic (EMI) interference, and may not show clear links between occupancy and sensor values. Past studies have indicated the need for a heterogeneous multi-sensory fusion approach for occupancy detection to address the short-comings of existing occupancy detection systems.

The aim of this research is to develop an advanced instrumentation strategy to monitor occupancy levels in non-domestic buildings, whilst facilitating the lowering of energy use and also maintaining an acceptable indoor climate. Accordingly, a novel multi-sensor based approach for occupancy detection in open-plan office spaces is proposed. The approach combined information from various low-cost and non-intrusive indoor environmental sensors, with the aim to merge advantages of various sensors, whilst minimising their weaknesses. The proposed approach offered the potential for explicit information indicating occupancy levels to be captured.

The proposed occupancy monitoring strategy has two main components; hardware system implementation and data processing. The hardware system implementation included a custom made sound sensor and refinement of CO_2 sensors for EMI mitigation. Two test beds were designed and implemented for supporting the research studies, including proof-of-concept, and experimental studies. Data

processing was carried out in several stages with the ultimate goal being to detect occupancy levels. Firstly, interested features were extracted from all sensory data collected, and then a symmetrical uncertainty analysis was applied to determine the predictive strength of individual sensor features. Thirdly, a candidate features subset was determined using a genetic based search. Finally, a back-propagation neural network model was adopted to fuse candidate multi-sensory features for estimation of occupancy levels.

Several test cases were implemented to demonstrate and evaluate the effectiveness and feasibility of the proposed occupancy detection approach. Results have shown the potential of the proposed heterogeneous multi-sensor fusion based approach as an advanced strategy for the development of reliable occupancy detection systems in open-plan office buildings, which can be capable of facilitating improved control of building services. In summary, the proposed approach has the potential to: (1) Detect occupancy levels with an accuracy reaching 84.59% during occupied instances (2) capable of maintaining average occupancy detection accuracy of 61.01%, in the event of sensor failure or drop-off (such as CO_2 sensors drop-off), (3) capable of utilising just sound and motion sensors for occupancy levels monitoring in a naturally ventilated space, (4) capable of facilitating potential daily energy savings reaching 53%, if implemented for occupancy-driven ventilation control.

TABLES OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

AF_DIFF	Area under for the curve applied for FDIFF data instances
ANFIS	Adaptive neural-fuzzy inference system
AMR	Automated meter reading
AS_DIFF	Area under for the curve applied for SDIFF data instances
ASHRAE	American society of heating, refrigerating and air-conditioning engineers
AVR	Average of sensor measurement
BEMS	Building energy management system
CAS	Case temperature sensor notation
CFS	Correlation feature selector
CIBSE	Chartered institution of building services engineers
CI	Computational intelligence
CO_2	Carbon (IV) oxide sensor notation/levels
CT	Current transformers
DCV	Demand controlled ventilation
DSM	Demand side management
EMI	Electromagnetic interference
FDIFF	First order difference
FFT	Fast Fourier transform
FLS	Fuzzy logic system
IAQ	Indoor air quality
IESD	Institute of energy and sustainable development
IC	Integrated circuit
ICE	Intelligent control of energy

IR	Infrared
GA	Genetic algorithm
HMM	Hidden Markov model
HVAC	Heating, ventilation and air-conditioning system
JDL	The data fusion sub-panel of the joint directors of laboratories
LR	Linear regression model
MAPE	Mean absolute percentage error
MF	Membership function
NDIR	Non-dispersive infrared
NRSME	Normalised root mean square error
NN	Neural network
NTC	Negative temperature coefficient
PC	Personal computers
PCB	Printed circuit board
PI	Proportional-integral
PID	Proportional-integral-derivative
PIR	Passive infrared sensor
PTC	Positive temperature coefficient
R^2	Coefficient of determination
RAE	Relative absolute error
RBF	Radial basis function network
RF	Radio frequency
RFID	Radio frequency identification
RH	Relative humidity

RMSE	Root mean square error
RTD	Resistance temperature detector
SDIFF	Second order difference
SMPS	Switch mode power supplies
SND	Sound sensor notation
SU	Symmetrical uncertainty evaluation
SVM	Support vector machine
TEMP	Ambient temperature sensor notation
THI_LO	Total number of state changes from HIGH to LOW
TLO_HI	Total number of state changes from LOW to HIGH
TOS	Total occupied times
TOTAL	Total event observed for sound or PIR measurements
TONP	Total number of state changes
TVOC	Total volatile organic compound
VAR	Variance of sensor measurements
VAV	Variable air volume
VTHI_LO	Variance of THI_LO measurements
VTLO_HI	Variance of TLO_HI measurements
VTONP	Variance of TONP measurements
VOS	Variance of TOS measurements
VOC	Volatile organic compound
WEKA	Waikato environment for knowledge analysis data mining tool

NOMENCLATURE

ρ	Linear or Pearson correlation coefficient
σ_y	The standard deviation of the training outputs
ρ_a	Density of air
AIC	Akaike information criterion
B	Amplifier output
C_{pa}	Specific heat capacity of air
C_C	Coupling capacitor
C_i	Fan coefficient values at different fan speeds, *i*
C_t	Time constant capacitor
C	Class label
D	Virtual earth
ε	Insensitive loss function
E_H	Energy consumed for heating the air mass
E_V	Energy consumed running the ventilation fan
E	Total energy necessary for optimal ventilation
f_{flow}	Flow fraction or part-load ratio
f_{pl}	Part-load factor
f	Operating frequency
F_i	A feature
F	A full set of features

G	Gain of amplifier
$H(Y\|X)$	Conditional entropy y_i given x_j.
I	Current
k	Number of the number of free parameters in the model
L	Maximised value of the likelihood function for the estimated model
$\widehat{m_b}$	Number of occupants estimated from sensor measurements
m_b	Actual occupancy number in the test area observed the infrared camera
m_e	Vector of occupancy estimations from the deployed sensors
m_o	Vector of observations by the infrared camera
$Merit_S$	Heuristic "merit" of a feature subset S containing k features
n	Total data instances
$p(y_i)$	Prior probabilities for all values of Y
$p(y_i\|x_j)$	Conditional probability of y_i given x_j.
P_{MAX}	Electrical power used by the fan
P	Conditional probability of the class label
Q_E	Model estimations
$\bar{Q}_O$	Mean of the actual occupancy data
$Q_{o(MAX)}$	Maximum value of Q_O
$Q_{o(MIN)}$	Minimum value of Q_O
Q_O	Actual occupancy data
$\overline{r_{cf}}$	Mean feature class correlation
$\overline{r_{ff}}$	Average feature inter-correlation
R_d	Voltage divider resistor

R_f	Feed-back resistor
R_i	Input resistor
R	Regularization parameter
SS_{res}	Total sum of squared errors
S_i	The set F with the F_i removed from F
SS_{tot}	Total sum of squares
SR	Slew rate
τ	Transpose
t_o	Fresh outdoor air temperature
t_r	Test area internal temperature
t	Time in minutes
T	Time constant
U	Number of sensors
$\dot{V}$	Ventilation or fresh air rates
V_i	Input voltage
V_L	Space volume
V_o	Output voltage
$\dot{V}_{design}$	Airflow rate through the fan
$\overline{x_i}$	Mean of X values
X, Y	Pair of random variables
$\overline{y_i}$	Mean of Y values

PUBLICATIONS

The following are a list of peer-reviewed publications through the course of this research;

Brown, N., Bull, R., Faruk, F., Ekwevugbe, T., (2011) “Novel instrumentation for monitoring after-hours electricity consumption of electrical equipment, and some potential savings from a switch-off campaign”, *Energy and Buildings*, 47, pp.74-83.

Ekwevugbe, T., Brown, N., Fan, D., (2012) " A design model for building occupancy detection using sensor fusion," *In: Proceedings of the 6th IEEE International Conference Digital Ecosystem Technologies – Complex Environment Engineering, IEEE DEST-CEE 2012*, Italy, June 18th -20th , 2012.

Ekwevugbe, T., Brown, N., Fan, D., (2012) " Using indoor climatic measurements for occupancy monitoring" *In: Proceedings of the 3rd West Africa Built Environment Research Conference* , *WABER 2012* , Abuja, Nigeria, July 24th -26th , 2012.

Ekwevugbe, T., Brown, N., Pakka, V., Fan, D., (2013) “Real-time building occupancy sensing using neural network based sensor network" *In: Proceedings 7th IEEE International Conference on Digital Ecosystems and Technologies, DEST 2013*, Menlo Park, CA, United States, July 24th -26th , 2013.

Ekwevugbe, T., Brown, N., Pakka, V., " Real-time building occupancy sensing for supporting demand driven HVAC operations," *In: Proceedings of the 13th International Conference for Enhanced Building Operations, ICEBO 2013*, Montreal, Canada, October 8th -11th , 2013.

CHAPTER 1

INTRODUCTION

1.0 Research background

Global warming is one of the most disturbing concerns facing humanity today due to accelerated release of carbon dioxide (CO_2) and other greenhouse gases into the atmosphere as a result of human activities (IPCC, 2001). Several initiatives have been introduced to mitigate climate change, one of such is the Kyoto Protocol (which is a part of the United Nations Framework Convention on Climate Change), with the primary aim of reducing global carbon emissions. Kyoto Protocol sets binding targets for 37 industrialized countries and the European Union to reduce their carbon emissions to an average of 5% against 1990 levels (IPCC, 2001).

The UK is an active force in the fight to mitigate climate change having signed up to the legally binding Kyoto Protocol. Based on recommendations from the Royal Commission on Environmental Pollution (RCEP, 2000), the UK government adopted a target to reduce CO_2 emissions by 60%, of the then current levels [1990's], by 2050. Achieving this target is a key policy goal for the government (DTI, 2002), and energy efficiency in buildings is expected to play a significant role in achieving this objective, as described by DEFRA (2006).

Worldwide building energy use is expected to grow 45% over the next 20 years (WBCSD, 2008) and buildings use up to 40% of total energy use in many countries including the UK (WBCSD, 2008), (DEFRA, 2006). But a significant part of this energy is wasted in servicing unoccupied buildings: in the UK for example, up to 23-30% of the non-domestic service sector electricity demand can be from unoccupied lighting (Brown, 2010), Meyers et al. (2010), found 39% of US domestic building energy wasted due to unoccupied heating and cooling. With roughly about half of the energy used in buildings attributed to Heating, Ventilation, and Air conditioning (HVAC) systems (Pérez-Lombard et al., 2008), there is clearly great potential for energy saving through improved building operations. Therefore, it is crucial to investigate different means for energy reduction and management in buildings.

1.1 Research motivation

With the drive towards low-energy buildings, building energy management systems (BEMS) are often employed to reduce operational energy consumed in many non-domestic buildings. A building management system is an automated system used for monitoring and controlling HVAC as well as lighting operations in modern buildings. Advancements in sensor and telecommunication technologies have seen installation costs of BEMS reduce drastically, increasing its widespread use (Loveday and Virk, 1992, CIBSE, 2009). With installed BEMS, energy savings of up to 15% can be achieved (CIBSE, 2009). However, BEMS have failed to fully optimize energy consumption in many non-domestic buildings (Zeiler et al., 2006). BEMS sensors have been reported to suffer from long term drift, lack of scheduled maintenance and even outright technical failure (Levermore, 2000). Also, control strategies used for running BEMS may not always be optimal (Erickson et al., 2011). To compound this, conventional HVAC operations just make use of temperature and humidity as sole inputs for system control, which often leads to energy waste (Agarwal et al., 2010). For building control operations, it is challenging to find the balance between energy efficiency and a comfortable climate (Zhu et al., 2010). One possible solution to achieving energy efficiency in buildings is to couple occupancy information to control strategies, such that services are provided only when needed (during occupied times). Previous studies have proposed up to 56% HVAC related energy savings with the application of occupancy-driven HVAC operations (Sun et al., 2011, Tachwali et al., 2007).

Occupancy information can be considered as the number of persons in a building space, and the resulting activities from occupants being present (i.e. associated electrical and HVAC loads) (Li et al., 2012). Ideally, building controls should automatically respond to dynamic occupancy loads. However, current BEMS often lack this capacity and usually rely on fixed assumptions to operate HVAC and electrical systems, leading to possible energy waste. Accurate knowledge of localised real-time occupancy information can have compelling implications for building controls, which may enable energy savings, whilst maintaining a comfortable environment. For example, such information is useful for determination of HVAC heat loads (Chenda and Barooah, 2010), as well as optimal run time, required heating, cooling and distribution of conditioned air, and optimal selection of temperature set points (Li et al., 2012).

A precise and reliable measurement of occupancy still remains difficult. Current technologies have certain shortcomings, including sensor drift, privacy concerns, low quality parts, intrusiveness, change of use and insufficient commissioning. More reliable and robust building occupancy sensors can be produced using sensor fusion techniques, aiming to estimate occupancy levels by merging information from various indoor environmental sensors. Sensor fusion aims to merge the strong qualities of various sensors, whilst minimising their weaknesses, thus providing better performance, which may not be possible from a single sensor type. Replacing existing sensors with a couple of more reliable low-cost and non-invasive sensors, such as those developed from a sensor fusion process may offer potential cost reduction for building occupancy monitoring (Dodier et al., 2006). While this is the case, there is a shortage of any systematic methodology for developing robust and reliable occupancy monitoring systems from multi-sensory sources (Hutchins et al., 2007). Thus, there is a need to produce a methodology for robust occupancy estimation in non-domestic buildings using a multi-sensor fusion approach.

1.2 Research hypothesis

The main hypothesis studied in this work is;

The combination of information derived from low-cost and non-intrusive indoor environmental sensors using machine learning techniques can provide reliable occupancy estimations in a naturally ventilated open-plan building.

Chapters four and five demonstrate that a novel machine learning based data processing methodology can indeed provide reliable occupancy estimations in an open-plan office using a network of low-cost and non-intrusive multi-modal sensors.

1.3 Research aim

The aim of this work is to develop advanced instrumentation strategies to monitor phenomena such as occupancy levels in non-domestic buildings, whilst facilitating the lowering of energy use and also maintaining acceptable indoor climate.

The main research question that arises is:

What indoor environmental variables are relevant for development of a robust system for occupancy detection, with a view to reduce energy use in a naturally ventilated building using a sensor fusion approach?

1.4 Research objectives

1. To investigate and develop new techniques of non-invasive instrumentation systems for non-domestic buildings.
2. To investigate a novel technique for data fusion processing.
3. To develop a novel occupancy detection system for non-domestic buildings.
4. To investigate any relationship between various building variables such as indoor climate, energy (electricity) and occupancy.

1.5 Thesis outline

The remainder of this thesis is structured as follows:

Chapter two provides background information and a literature review on building occupancy detection systems, building instrumentation technologies, computational intelligence (CI) and its applications for building indoor environmental control and energy management, and multi-sensor data fusion. A detailed state-of-the-art survey on recent advances is given together with critical analysis and discussion.

Chapter three presents a detailed description of an advanced multisensory building occupancy instrumentation strategy. First, it presents the experimental design used in the research. Next, findings from a pilot experiment are discussed. Finally, modification process for CO_2 sensors, and the custom hardware design and implementation for sound sensing are presented.

Chapter four describes a novel data processing methodology for building occupancy detection. The data processing system comprises of several stages, namely, pre-processing, feature ranking, feature selection and feature fusion for occupancy estimation. Experimental test results are presented.

Chapter five presents experimental results for the application of individual variable sensing network, and a heterogeneous multi-sensor network for occupancy estimation. The issue of resilience in occupancy sensing network, and cross room model analysis are also presented.

Chapter six explores the relationship between different building variables such as occupancy, indoor climate, and energy (electricity) use. Potential energy savings using occupancy information produced by the proposed detection system is confirmed by experimental test.

Chapter seven provides the discussion and conclusion of the research. The main contributions to knowledge and the impact of this study to a range of stakeholders within the built environment are discussed in detail. Recommendations are given for future research.

CHAPTER 2

LITERATURE REVIEW- BUILDING INDOOR ENVIRONMENTAL MONITORING AND CONTROL

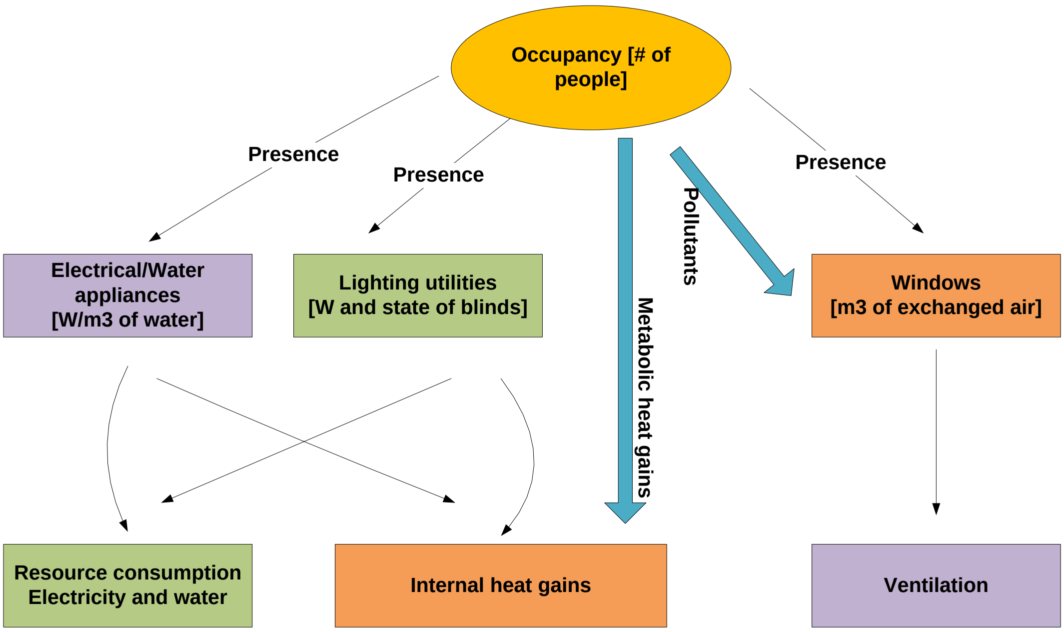
2.0 Introduction

This chapter discusses background information on building environmental monitoring and control. It also presents a critical discussion of existing research in related fields. The interdisciplinary nature of this research requires a comprehensive review of literature related to building occupancy detection methodologies, building instrumentation technologies, computational intelligence (CI) and its applications for building indoor environmental control and energy management and multi-sensor data fusion. The primary aim of this chapter is to provide an in-depth description of common building sensing technologies, control strategies, and also to identify useful data processing techniques for occupancy detection systems.

This chapter is organised as follows: section 2.1 provides a detailed account of occupancy detection approaches in the literature. Section 2.2 introduces common building instrumentation technologies and their principles of operation. Section 2.3 discusses various commonly used computational intelligence techniques, while section 2.4 provides insights on state-of-the-art applications for indoor comfort control. Section 2.5 explores opportunities for energy savings using occupancy-driven power management of office appliances, and implementation of a range of demand-driven HVAC control strategies. Section 2.6 summarizes the concept of sensor fusion, virtual sensing and their potential for building instrumentation. Section 2.7 presents a summary of the findings from the literature review.

2.1 Building occupancy detection

Occupants' presence and behaviour impacts on HVAC related energy use, as well as that of electrical appliances. Occupants' interactions with their indoor environment can be useful proxies for developing occupancy detection systems. They affect the indoor environmental conditions through the release of body heat, carbon-dioxide, odour, water vapour, and sound as a result of their activities. Figure (2.1) illustrates the interaction of occupants and their environment. Building occupancy detection using low-cost, non-invasive environmental sensors may be useful for occupancy driven HVAC operations, yet relatively limited research has been conducted in this field especially in office buildings.



Adapted from (Page et al., 2008)
Figure 2.1: Occupants' interaction with their environment.

The impact of occupancy numbers on building energy use can be explored using models such as the American Society of Heating, Refrigerating, and Air-Conditioning Engineers (ASHRAE) Standard 90.1 (which defines several occupancy profiles for office building day types) (ASHRAE, 2007), especially during the design phase using simulation tools (e.g. Energy Plus). However, there is a growing interest in the development of occupancy detection systems that combine raw sensor data with advanced algorithms. Various occupancy detection systems in the literature often indicate overlapping research and development in many applications, as it may be sometimes difficult to make a clear distinction between these systems. A review of related works is therefore discussed under two broad headings; namely simulation and sensor-fusion based approaches.

2.1.1 Simulation approach

Previous studies in this area have attempted to describe occupancy dynamics in a building using mathematical models, some of which were applied in simulation softwares. Several stochastic models have been proposed for modelling occupants' presence and interactions with their space. Reinhart (2004) developed a simplified stochastic model of occupants' arrival and departure as an input for their Lightswitch-2002 algorithm. A Monte-Carlo modelling approach was used by Degelman (1999) to predict occupancy profiles in an office building; the model was based on survey statistical data on how people use office spaces. Richardson et al. (2008) proposed a stochastic occupancy model for UK households based on 10-minute resolution binary data. Page et al. (2008) developed a stochastic occupant presence model, which was considered as an inhomogeneous Markov chain interrupted by occasional periods of long absence. A time series model of each occupant in a single person office was generated, with each having a different profile describing his/her probabilities of presence in the office at different times of the day. Simulated model results matched well with the original motion sensor data used to develop the profile, in terms of occupants' probability of presence. However, model accuracy for occupancy presence prediction was not provided, rather it only produced a probability density for occupancy at that day and time. It was not clear if the model would be effective for occupancy presence prediction in multi-occupant spaces (e.g laboratories and conference rooms).

The effectiveness of Page's model (Page et al., 2008) is also weakened by findings from Wang et al. (2005), who studied occupancy pattern in 35 single person offices, and found out that vacancy intervals followed an exponential distribution while occupancy interval distribution was time varying. The basic assumption used in the study was that a room is in a state of being vacant to occupied when passive infrared (PIR) sensors detected motion, and occupied to vacant when the sensors detected no motion within a time interval of 15 minutes. This assumption may not be a sufficient condition to classify the room as vacant, since it is possible that occupants can make minor movements or remain stationary for long periods inside the office, which a PIR sensor may fail to detect (Tiller et al., 2010).

Occupancy has also been modelled with linear regression, using lighting and equipment load data, Abushakra and Claridge (2008), developed linear regression models to show correlation between occupancy and electrical loads during weekdays, weekend and holidays. Ground truth occupancy data were determined from walk-through surveys. The main drawback of this model is that it relies on energy usage for determination of occupancy, which is a problem because this model tends to underestimate occupancy. For instance, a large number of occupants using a lecture theatre may not necessarily increase the room's electrical load, causing the model to report the room as vacant.

Others have used computational intelligence (CI) based models, Chenda and Barooah (2010) and Liao et al. (2012) estimated occupancy number in an office by fusing sensor data with predictions from a complex agent-based stochastic model. Simulation of occupants' behaviour was carried out using a mixed agent-based rules model, and a graphical model was used to establish probabilistic factors that affect agent behaviour. It is not clear how the model would perform for multi-occupant spaces, since the experiment was limited to single occupant office, and besides the graphical model cannot by itself predict occupancy due to the uncertain nature of occupants' activities. Based on PIR sensor data, Yu (2010) applied a genetic algorithm to learn the behaviour of an occupant in a single person office. Prediction accuracy of 80-83% was achieved, although there were no ground truth data to validate model performance.

Generally, occupancy models in this category tend to be applied to single occupants' spaces, where occupancy dynamics is relatively simple. It was not clear, how these models can be applied in non-domestic buildings where the occupancy dynamics can be a more complex phenomenon. Figure (2.2) summarises some of the techniques applied in modelling occupancy under the simulation approach.

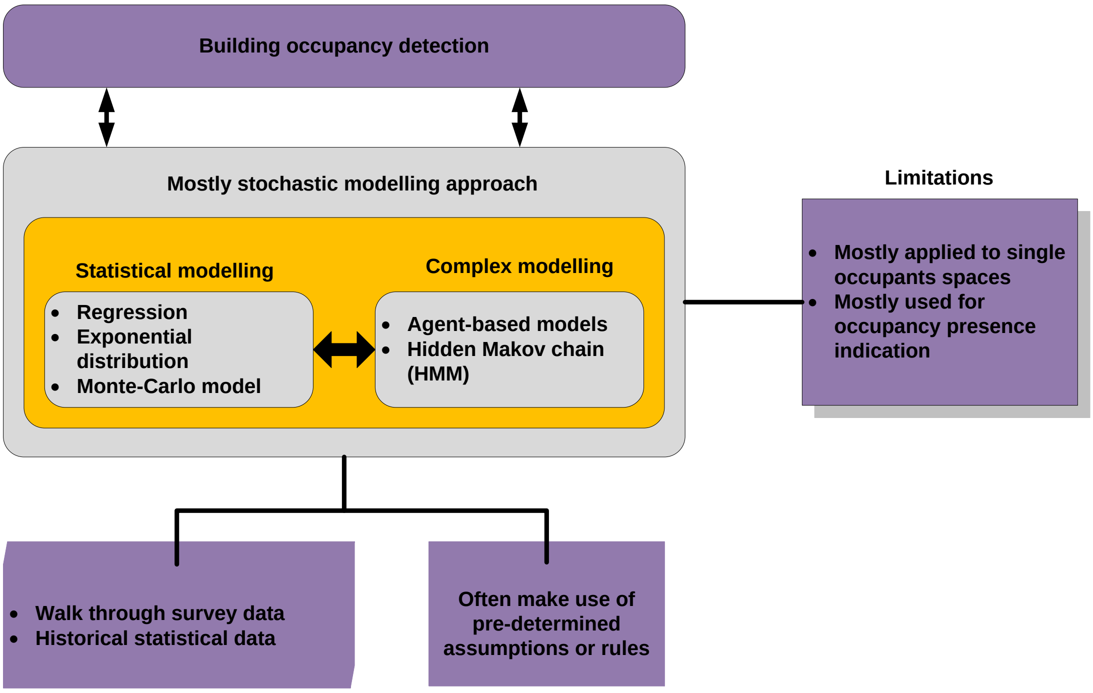


Figure (2.2): Simulation approach

2.1.2 Sensor fusion based approach

Using this approach, physical measurements taken from an indoor environment are analysed using sophisticated algorithms to extract occupancy information. Some occupancy detection systems make use of redundant sensors while others use a combination of different ambient sensors. Non-ambient sensors such as wearable sensors have also been considered under this category for this research.

- **Wearable sensors**

A limited number of studies have reported the use of wearable sensors for occupancy monitoring in office buildings. Devices could be in the form of a ring (Sokwoo et al., 1998) or wristwatch (Lötjönen et al., 2003), or neck tag. The use of these devices is widespread in health care monitoring. Korhonen et al. (2003) proposed the use of wearable sensors for monitoring recovering hospital patients, while Sungmee and Jayaraman (2003) suggested the use of a smart shirt for monitoring health conditions (such as heart beat rate, temperature etc.) of its users. This shirt may be useful in a heath care setting, but it is clearly impractical for use in other places such as public buildings.

Radio frequency identification (RFID) based sensors can be effective for indoor occupancy monitoring. For example, Li et al. (2012) proposed an occupancy detection system based on RFID tags, which reported real-time occupancy numbers and the thermal zones where each occupant was located in an office building. A K-nearest neighbour algorithm was used for occupancy tracking, and the system produced an average zone detection accuracy of 88% for stationary occupants, 62% for moving occupants. Zi-Ning (2008) developed an occupancy detection system for lighting control. Occupants' localization was achieved using a support vector machine (SVM) algorithm that was aided by a round-robin rule based on some numerical logics. Tracking accuracy reached 93% for occupants' that wore the RFID tags. However, it was not clear how the system would address issues of latency and scalability in large non-domestic buildings with large occupancy profiles. Gillott et al. (2009) and Gillott et al. (2010) determined occupancy patterns in a residential building using ultra-wideband RF-based tags worn by occupants, the system tracked moving occupants to within an accuracy of 15cm in three dimensions. One advantage of these technologies is that they provide occupancy information that can be based on zones that are either physically or virtually partitioned, making them suitable for use in open-plan spaces with multiple thermal zones (Li et al., 2012). However, willingness and ability of occupants to wear these devices may be a critical issue for their uptake.

- **Vision-based systems**

Vision-based systems are mostly deployed for security and access control in buildings. Chen et al. (2006) proposed a video-based system using area and colour information to monitor people flow through a gate. Pedestrian patterns were extracted from raw images using motion analysis, while pedestrian direction was recognised with hue, saturation and intensity (HSI) analysis. Accuracy was improved by coupling pedestrian's colour information to the earlier counting output. Results showed 100% accuracy for bi-directional people counting; when the people number of a people-touching pattern was less than six. The system may be useful for building access control, but it remains unclear how it can be applied for occupancy estimation in building zones with numerous exits such as lecture theatres.

Vision-based systems, which rely on camera images and image analysis techniques are currently been experimented for occupancy detection in buildings. For instance, Benezeth et al. (2011) developed a sensor for detection of human presence and characterization of their activity in an office using a network of video cameras. Image analysis was divided into three stages- change detection, moving objects tracking and classification. The first stage used a background subtraction algorithm to detect only what changed in the environment, and in the second stage, features of interest from the objects were tracked. Finally, the nature of the object was determined, to classify whether it was human or not, using a multiple cascade of boosted classifiers. Detection rate for the number of occupants inside offices reached 93% and 83% for corridors. Tomastik et al. (2008) estimated occupancy in different zones in a building using video cameras with an inherent signal processing algorithm that detected the number of people passing through its field of view. This occupancy estimator used an extended Kalman filter based on a non-linear stochastic state space model of people's traffic and the video's sensing output to infer occupancy in a zone. Experiments produced better estimation performance than using video cameras alone. This approach holds potential for estimating occupancy in building zones that are not completely observable by installed occupancy sensors.

İçoğlu and Mahdavi (2007) described a self-updating vision-based model for object identification and location sensing as well as occupancy detection in sentient buildings. The model processed images from cameras with a distributed data fusion algorithm for occupancy sensing. It offers flexible integration and scalability for components (hardware and software) of an existing building control system, such that communication sharing was internet driven. One of the drawbacks of the model is that it requires large parallel computational resources, in cases where the deployment area is an entire building with several rooms. Sarkar et al. (2008) used standard type of video cameras to monitor occupancy. The system was only useful for occupancy presence and vacancy status indication, which clearly limits its application. Erickson et al. (2009), Erickson et al. (2011) proposed the use of wireless camera network for coarse-grain occupancy detection at the floor level. Occupancy information was modelled as an inhomogeneous Markov chain using images collected from the camera network, and detection accuracy reached 71.20%. While an annual HVAC energy saving of 42% was achieved using control strategies based on the system occupancy information, substantial installation and maintenance overhead costs may be a concern. Silvestre and Perez (2011) proposed a system using multimedia technology for occupancy –driven lighting control in an industrial building. This ran on common image processing algorithms, which carried out processes such as background subtraction, blob tracking and analysis, to detect occupants' presence and movement. An estimated energy saving of 70% was reported, by coupling the system to existing lighting controls.

Despite the high detection rate of some vision-based systems, privacy concern remains a major factor limiting their implementation. Algorithms currently being developed to address this problem are still in the early stages of commercial maturity. Vision-based systems suffer a common drawback which requires them to have a clear line of sight in the observed spaces, which clearly limits their applicability in heavily partitioned spaces. Besides, video cameras also require large image storage facilities. Although, some smart cameras may not require this.

- **Carbon-dioxide based systems**

Carbon-dioxide (CO_2) sensors are widely used for regulating CO_2 levels especially for demand controlled ventilation (DCV) systems (Nielsen and Drivsholm, 2010, Kar and Varshney, 2009, Sun et al., 2011). This may be unsuitable for conditioning strategies, as CO_2 build-up is often slow, such that by the time sensors detect high levels of CO_2 that trigger ventilation, occupants may already be in state of discomfort (Fisk, 2008). Some works have attempted to improve the reliability of CO_2 based occupancy sensing. For example, Dong et al. (Dong et al., 2010, Lam et al., 2009b, Lam et al., 2009a) proposed a system that used information from CO_2, acoustic and PIR sensors to estimate the number of occupants in an open-plan office space. Using information theory, the most relevant information for occupancy prediction was extracted from sensor data, and fused with three machine learning algorithms (support vector machine, artificial neural networks, and hidden Markov model). An average reported accuracy of 73% was achieved by the hidden Markov model. Meyn et al. (2009) improved occupancy detection accuracy by using a sensor network comprising CO_2 sensors, digital video cameras and PIR detectors as well as historic building utilization data for occupancy estimation at the building level. The system used a receding-horizon convex optimization algorithm to infer occupancy numbers. System accuracy reached 89% for the entire building; however it was not able to estimate occupancy numbers at the room level.

Past studies have also investigated the use of CO_2-based occupancy detection systems in residential buildings, as opposed to a mixed-use building. For example, Cleveland and Schuh (2010) developed an occupancy monitoring system for automation of HVAC thermostats in residential buildings using CO_2 and motion sensors, and a simple control algorithm based on the rate of change of CO_2 levels. Occupancy detection was best inferred from CO_2 levels in the house. CO_2 levels of 525ppm or a change in CO_2 concentration reaching 50ppm or above for two straight minutes, indicated occupancy, while concentration of 300ppm suggested vacancy. Results were not validated with a field test, and it remains unclear how the system would perform for occupancy number estimation.

CO_2-based systems may be susceptible to common operational limitations, since they generally have slow response in detecting incoming people (Wang and Jin, 1998), and also CO_2 concentration levels may be affected by factors other than occupancy such as passive ventilation (e.g. open windows, air infiltration etc.). These sensors may suffer significant drift over time (say over a year) which may limit their functionality (Shrestha and Maxwell, 2010). Besides, there is a high level uncertainty between number of persons and CO_2 concentration levels (Chenda and Barooah, 2010). Such limitations make accurate and robust prediction of real-time occupancy numbers using CO_2-based systems challenging.

- **Other ambient sensors**

Passive infrared sensors are the most commonly used technology for occupancy sensing in non-domestic buildings especially for lighting control (Delaney et al., 2009), however they fail to detect stationary occupants, thus switching off services falsely. Garg and Bansal (2000) proposed a smart occupancy sensor that adapts to changing activity levels of occupants in a building space. The authors demonstrated that by varying a PIR sensor time delay with respect to a known activity pattern of an occupant, the number of false-off's can be minimised. However, in cases where occupancy patterns are uncertain, variation in time delay alone may not completely eliminate the problem of false-offs. To address this problem, PIR sensors are coupled with other sensors. Dodier et al. (2006) proposed a Bayesian belief network, comprising of three PIR sensors and a telephone sensor to probabilistically infer occupancy. Occupied state of individual offices room was modelled with a Markov chain. Their system had a detection accuracy of 76%, but was unable to count the number of occupants. Hailemariam et al. (2011) deployed a sensing test-bed, consisting of CO_2, light, sound and motion sensors for detection of occupancy presence or absence in office cubicles. Sensor data were combined with a decision tree algorithm. A reported detection accuracy of 98.4% detection was achieved when PIR sensors were used alone. However, accuracy decreased when other sensors were added to the fusion process. In addition, the system is limited for occupancy numbers estimation. Padmanabh et al. (2009) used a combination of microphones and PIR sensors to gather occupancy information for efficient scheduling of a conference room. The room was considered occupied (meeting ongoing) if a microphone value exceeded a threshold twice in a 5-minute interval; otherwise it was classified as unoccupied (no meeting).

A combination of door (reed switch) and PIR sensors have been utilized for occupancy presence detection in buildings (Lu et al., 2010), (Agarwal et al., 2011), (Agarwal et al., 2010). In Lu et al. (2010), room occupancy status was inferred from sensor data by means of a probabilistic model based on a hidden Markov chain, which classified occupancy based on daily occupant schedules in to one of the three states based: occupied with all occupants asleep, occupied with an occupant awake, and unoccupied. A classification accuracy of 88% for occupied instances was reported, although it remains unclear how the model would perform given a different daily schedule for the occupants.

In an attempt to improve the robustness of multi-sensory occupancy detection systems, Hutchins et al. (2007) proposed an approach that used data from two optical people counting sensors and a probabilistic model to estimate occupancy in a building. The sensor registered a count when the optical beam was interrupted. However, due to sensing imperfections such as over counting and under counting, the authors used a probabilistic model (consisting of an inhomogeneous Poisson process and a hidden Markov process) to estimate occupancy, taking into consideration measurement noise and historic data. Although, results suggested the approach could recover up to 50% missing sensor data, the system was not validated with field tests.

Other studies have explored occupancy detection in residential buildings within the context of home activity recognition and elderly care applications. Fogarty and Hudson (2006) used a network of low-cost sound sensors to monitor activities of occupants in a household. These were attached to the water pipes around the house and data collected were then processed with a support vector machine (SVM) algorithm to recognise occupants' activities. System accuracy reached 97% recognition for toilet flushing events. Jianfeng et al. (2005) built a novel bathroom activity recognition system consisting of microphone and PIR sensors, which was able to detect and recognise real-time sound events with a hidden Markov model to an accuracy of 87%. Both systems highlight the usefulness of sound sensing for activity monitoring, and hence occupancy presence detection. However, their functionality is prone to external interference which may limit their performance. Wilson and Atkeson (2005) applied a range of binary sensors such as motion sensors, beam sensors, pressure mats and contact switches for occupant tracking and activity recognition at the room level, although system performance is limited to whether or not an occupant is moving. Hong et al. (2009) improved the robustness of activity recognition systems by proposing evidential fusion networks running on a Dempster-Shafer algorithm comprising of light, sound, motion and contact sensors. The system was capable of accommodating uncertainties in sensor data, and identifying a threshold for the minimum number of sensors required to gather sufficient information for activity recognition. The main limitation of these systems is that they may not be appropriate, if the purpose of monitoring is to establish occupancy numbers.

- **ICT-based systems**

A number of research studies have highlighted the feasibility of occupancy detection in offices by monitoring electronic appliance usage. For example, Melfi et al. (2011) developed a novel occupancy detection system running on an existing IT infrastructure. The system monitored occupants' MAC and IP addresses, keyboard and mouse activities as occupancy proxies. Reported detection accuracy of 80% at the building level, and 40% at the floor level was achieved. Martani et al. (2012) studied the relationship between building occupancy and energy use using the number of Wi-Fi connections as proxies for occupancy estimation. Overall, at the building level, occupancy accounted for between 63% and 69% variation of the total electricity consumption. Brown et al. (2011), and Brown and Wright (2008) proposed a useful method for establishing the usage patterns of electronic appliances (such as desktop PCs), from which occupancy can also be inferred. Using portable temperature sensors attached to the case of PCs and a pinging software routine that ran on the local network, appliance duty cycles were detected to a precision in excess of 97%. However, the main disadvantage here is that these methods are not able to detect occupants that do not use PCs. Krumm et al. (2007) proposed a sensor approach for detection of electrical noise on residential power lines created by electrical appliances when turned on, where the authors exploit the fact that each appliance produces a unique noise signature. A machine learning algorithm was used to characterize an appliance from the electrical noise it generated when turned on or off to an accuracy of 85-90%. A similar approach was adopted by Kim et al. (2009), but made use of non-intrusive sensors such as light intensity sensors, magnetometers, and microphones to monitor the usage pattern of electrical home appliances. Christensen et al. (2004) reported early work in the use of internet traffic characterization to establish usage patterns of desktop PCs within a local network in a university dormitory. Idle time was considered as a period with no active internet traffic, while a busy period as a time of overlapped internet traffic. Energy savings of 1TWh/year in the US were proposed, if PCs were powered down during idle time.

Other studies focus on embedding occupant's physical detection capacity into computing systems. Tarzia et al. (2009) used inbuilt PC sonar alongside a threshold based algorithm to detect the presence of computer users. This technique relied on the principle that a human body produces different effects on sound waves, than air and other objects. Reported user detection accuracy reached 96%. The use of eye tracking for PC power management was suggested by Moshnyaga (2010). A video camera placed on top of the desktop monitor was used to track the user's eye, keeping the computer active only if the user was looking at the desktop monitor; otherwise, it was powered down. This approach may be limited in circumstances such as when more than one user is looking at the screen at a particular instance. Generally, ICT-based occupancy detection systems may be suitable for occupancy driven- power management of electrical appliances but limited for occupancy numbers estimation in buildings, as occupants not using computers (or other electrical appliances) are not detected. Figure (2.3) illustrates a multi-sensory fusion based approach for building occupancy detection. It presents a summary some of the techniques employed as described in the reviewed literature.

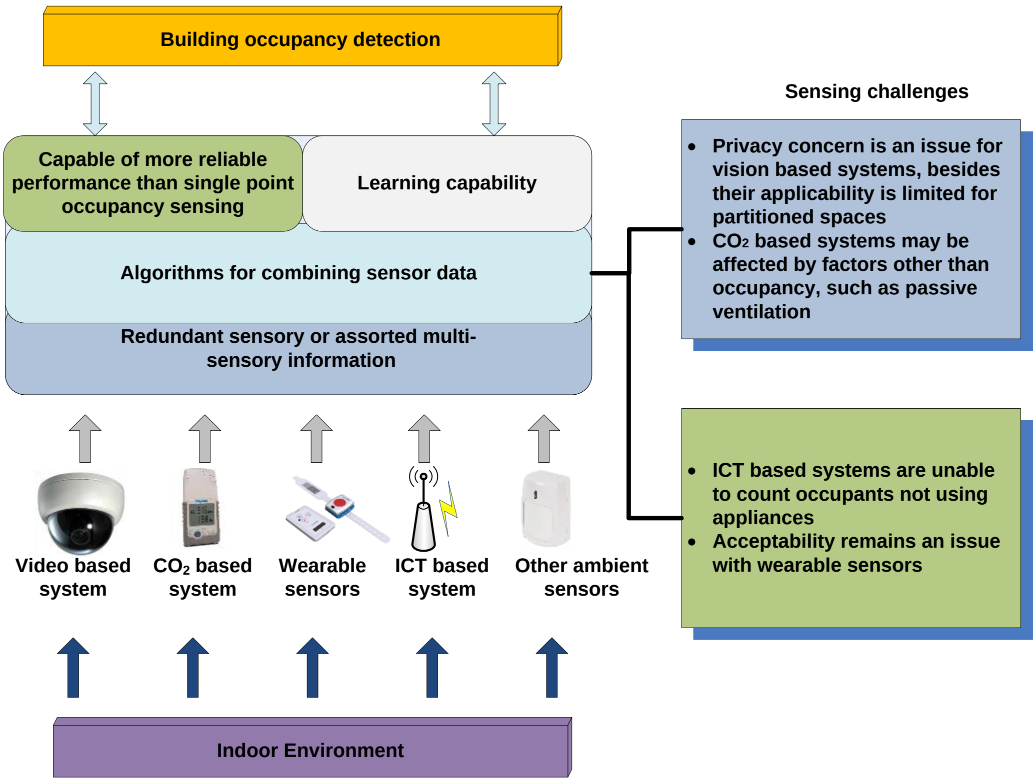


Figure (2.3): Sensor fusion approach

2.1.3 Discussion of building occupancy detection system

There is an increasing amount of literature on the need for occupancy-driven control of HVAC, lighting as well as other electrical systems in buildings for reducing energy use. Table 2.1 gives a summary of various occupancy detection systems in the literature. Most of the occupancy presence models in the literature were targeted at single occupant offices or residential buildings, where HVAC systems are generally less complex as opposed to that in non-domestic buildings. In commercial office settings (such as in university buildings), occupancy schedules are largely dynamic, with academics often travelling, teaching or attending meetings elsewhere, coupled with uncertain student occupancy patterns in classrooms. This is compounded by the complex nature of their HVAC systems which may have many interlinked controllers, and sensors for regulating indoor climatic conditions of several hundreds of zones within a building. It remains unclear how these models would perform in spaces with highly dynamic occupancy patterns.

In general, occupancy predictions using simulation approaches tend to be probability densities, rather than specific real-time predictions. The sensor fusion based approach attempts to address this problem, although a number of occupancy detection systems in the reviewed literature, have certain short-comings with respect to accuracy, cost, intrusiveness, and privacy. ICT-based systems cannot detect occupants not using electronic appliances; vision-based systems can do this but are limited by privacy concerns. Wearable sensors can also be used to obtain absolute real-time occupancy count but willingness of occupants to wear the devices is a major drawback. The need for a heterogeneous multi-sensory approach coupled with a machine learning algorithm for occupancy detection has been clearly advocated in the literature, with some system detection accuracy reaching up to 98%, but with obvious limitation for occupancy numbers estimation. Further research is recommended for development of improved and robust occupancy detection systems, and sensor fusion techniques offer a way to make up for this, aiming to determine occupancy levels more reliably using a range of different indoor climatic variables. This research adopts the use of low-cost and non-invasive multi-sensor network for occupancy detection.

Table (2.1): Summary of various building occupancy systems

Studies	Sensor used	Detection category	Features extracted & data collected	Detection algorithm	Detection accuracy	Limitations
(Korhonen et al., 2003), (Sungmee and Jayaraman, 2003)	Wrist-sensor, smart shirt	Activity monitoring in health setting	Heart rate, body temperature, respiration rate level etc.	Not clear from the study	Not clear from the study	Willingness for occupants to wear devices
(Li et al., 2012)	RFID tags	Occupancy numbers and localization	Not clear from the study	K-nearest neighbour algorithm	88% for stationary occupants, 62% for moving occupants	Willingness for occupants to wear devices
(Zi-Ning et al., 2008)	RFID tags	Occupancy tracking	Not clear from the study	Support vector machine (SVM)	93% for occupants wearing tag	Willingness for occupants to wear devices
(Chen et al., 2006)	Video sensing	Bi-directional people counting	Area and colour information	Hue ,Saturation and intensity (HSI) analysis	100% for bi-directional people counting	Privacy concerns
(Benezeth et al., 2011)	Video sensing	Occupancy numbers and tracking	The whole body (regardless of view point), The upper-body (head and shoulders) from front or back views, The upper-body (head and shoulders) from left view, and The upper-body (head and shoulders) from right view.	Haar filter algorithm	93% for occupants inside the office, and 83% for corridors	Privacy concerns
(Tomastik et al., 2008)	Video sensing	Occupancy numbers	Traffic model and motion information	Extended Kalman filter	Not clear from the study	Privacy concerns
(İçoğlu and Mahdavi, 2007)	Video sensing	Occupancy presence	Occupant images	Custom algorithm-distributed data fusion algorithm	Not clear from the study	Privacy concerns, limited for occupancy numbers estimation
(Sarkar et al., 2008)	Video sensing	Occupancy presence	Absolute YCC image difference the last frame and the current	Digital colour imaging algorithm, daylight sensing	Not clear from the study	Privacy concerns, limited for occupancy numbers estimation

			frame	algorithm		
(Erickson et al., 2009), (Erickson et al., 2011)	Video sensing	Occupancy numbers	Occupancy is partitioned in to hourly blocks	Hidden Markov chain (HMM)	System accuracy reached 71.2% for Lab occupancy detection	High installation cost, privacy concern
(Silvestre and Pérez, 2011)	Video sensing	Occupancy presence	Occupants images	Background subtraction and blob tracking algorithms	Not clear from the study	Privacy concerns, limited for occupancy numbers estimation
(Lam et al., 2009b), (Dong et al., 2010)	CO_2, PIR, and acoustic sensors	Occupancy numbers	Raw CO_2, motion, 20-minute moving average for CO_2 levels, first order shifted difference for CO_2 level, and first order difference for between outdoor and indoor CO_2 levels, acoustic levels.	Neural network, SVM, HMM	73% using HMM	CO_2 concentration may be affected by factors other than occupancy such as passive ventilation (e.g open widows, air infiltration etc)
(Meyn et al., 2009)	CO_2 , PIR, video camera, historic data	Occupancy numbers	Room utility information, CO_2 levels, motion	Receding –horizon convex optimization algorithm	89%	System can estimate occupancy at the room level
(Cleveland and Schuh, 2010)	CO_2 and PIR	Occupancy presence	Change in CO_2 concentration above 50ppm for a 2-minute interval for occupancy. 300ppm for vacancy	Simple threshold algorithm	Not clear from the study	Limited for occupancy numbers estimation.
(Dodier et al., 2006)	PIR and telephone hook	Occupancy presence	Probabilities of presence from sensory data	Bayesian belief algorithm	76% for presence detection	Limited for occupancy numbers estimation.
(Hailemariam et al., 2011)	CO_2, light, sound, and PIR	Occupancy presence	Average sensor reading, Root mean square (RMS) error	Decision tree	98.4% for presence detection	Limited for occupancy numbers estimation.
(Padmanabh et al., 2009)	Microphones and PIR	Occupancy presence	Microphone value exceeding a threshold twice in a 5-minute interval	Simple threshold algorithm	Increased the utility of conference room by 23%.	Interference from external sources, when the room are not sound proof
(Agarwal et al., 2010,	Contact sensor(reed	Occupancy presence	Door events (Open or	Simple threshold	Not clear from the	Limited for

Agarwal et al., 2011)	switch) and PIR		closed), motion	algorithm	study	occupancy numbers estimation.
(Lu et al., 2010)	Contact sensor (reed switch) and PIR	Occupancy presence	occupied with all occupants asleep, occupied with an occupant awake, and unoccupied	HMM	Reported an 88% classification accuracy for occupied instances	Limited for occupancy numbers estimation.
(Hutchins et al., 2007)	Optical sensors, historic data	Occupancy numbers	Occupants interrupting an optical beam	HMM	System could recover up to 50% missing data	System was not validated with field test
(Fogarty and Hudson, 2006)	Sound	Occupancy presence: activity recognition	Zero-crossing rates, and root mean square of microphone samples	SVM	System accuracy reached 97% recognition for toilet flushing event	Limited for occupancy numbers estimation.
(Jianfeng et al., 2005)	Microphone and PIR	Occupancy presence: activity recognition	Mel- Frequency Cepstral Coefficient (MFCC) of sound events	HMM	87% for presence detection	Limited for occupancy numbers estimation.
(Wilson and Atkeson, 2005)	PIR, beam sensors, pressure mats, contact switches	Occupancy presence :activity recognition and tracking	Binary events	Rao-Blackwellised particle filter	Detection accuracy reached 98% for a two person experiment, and 86.4% for a three person experiment	Limited for occupancy numbers estimation.
(Hong et al., 2009)	Light, sound, motion and contact sensors	Occupancy presence: Activity recognition	Binary events	Dempster-Shafer algorithm	Not clear from the study	Limited for occupancy numbers estimation.
(Melfi et al., 2011)	IT infrastructure	Occupancy presence	MAC and IP addresses, keyboard and mouse activities	Simple algorithm to count the number of hosts as indicated by Address Resolution Protocol (ARP) or Dynamic Host Control Protocol (DHCP) used by routers.	80% at the building level, and 40% at the floor level	It cannot detect occupants not using a computer
(Martani et al., 2012)	IT infrastructure	Occupancy presence	Wi-Fi connections	Regression analysis	Not clear from the	It cannot detect

					study	occupants not using a computer
(Brown et al., 2011)	IT infrastructure, temperature sensors	Occupancy presence	Change in case temperature, Ethernet network connections	Simple threshold algorithm	97% duty cycle detection	It cannot detect occupants not using a computer
(Krumm et al., 2007)	Power line noise analyser	Occupancy presence	Appliance noise signature : Drastic changes in the input line noise in 1μs	Fast Fourier Transform (FFT), SVM for classification	85-90% classification accuracy	It cannot detect occupants not using appliances
(Christensen et al., 2004)	IT infrastructure	Occupancy presence	Internet traffic: Busy and idle periods	Simple custom algorithm to determine active or idle computer	Not clear from the study	It cannot detect occupants not using a computer
(Tarzia et al., 2009)	PC sonar	Occupancy presence	Echo variance from emitted body sound waves	Custom state classifier algorithm based on threshold analysis	96% for presence detection	It cannot detect occupants not using a computer
(Moshnyaga, 2010)	PC video camera	Occupancy presence	Between-The-Eyes (BTE) pattern of a human face	SVM	84% detection accuracy	It is not clear how the system will perform when multiple occupants are in front of the computer , besides it cannot detect occupants not using a computer

2.2 Building instrumentation technologies

Advances in electronics have seen the cost of sensors falling, such that there are numerous sensors available for building monitoring. While this is a good thing, it may have also introduced some level of confusion when choosing sensors for building monitoring. This section presents state-of-the-art in building instrumentation technologies. Table 2.2 provides some of the factors considered in selecting a suitable sensor. The sensing technologies are described in terms of general operating principles, common uses, and limitations.

Table 2.2: Influencing factors for building instrumentation selection

Sensor Parameter	Description
Range	Difference between the maximum and minimum value of the sensed parameter
Resolution	The smallest change the sensor can differentiate
Accuracy	Difference between the measured value and the true value
Precision	Ability to reproduce repeatedly with a given accuracy
Sensitivity	Ratio of change to a unit change of the input
Zero drift	The departure from zero value over a period of time for no input
Response Time	The time lag between the input and output
Bandwidth	Frequency at which the output magnitude drops by 3dB or range of frequencies that are not inherently affected by the device
Operating temperature	The range in which the sensor performs as specified
Deadband	The range of input for which there is no output
Specificity or selectivity	The ability to detect a target gas without being affected by the presence of interfering gases
Repeatability	Closeness of the agreement between the results of successive measurements of the same measurand carried out under the same conditions of measurement. Repeatability can be assessed when the sensors are subject to precisely calibrated gases samples
Reproducibility	Closeness of the agreement between the results of the measurements of the same measurand carried out under changed conditions of measurement.
Hysteresis	The difference in response of the sensor when calibrating from a zero to mid-scale compared to the response when calibrating from full scale to mid-scale

Source : (Bishop, 2002)

2.2.1 Building occupancy sensing

Occupancy sensors act as switching devices that respond to occupants' presence/absence within their field-of-view. Several building occupancy instrumentations are available, although two technologies dominate; PIR and ultrasonic sensors. PIR sensors are commonly used for building occupancy sensing especially in lighting control applications (Delaney et al., 2009). The main sensor components in a PIR sensor are a pyroelectric detector and a lens. They are most sensitive to moving objects that emit heat energy at around 10μm (Moghavvemi and Seng, 2004). When a PIR sensor detects temperature changes within its field of view, the pyroelectric material undergoes a change in polarisation which produces a voltage signal. Its sensitivity decreases when distance between the sensor and a moving warm object increases (Kaushik and Celler, 2007). PIR sensors tend to work well where the entire observed space is within their direct line-of-sight, although they fail to detect stationary occupant (Benezeth et al., 2011). Consequently, they may switch off services in occupied spaces causing inconvenience to occupants. PIR sensors are available as an off-the-shelf product, with the cheapest one sold for £7.

An ultrasonic sensor uses high frequency sound (between 25 and 40 kHz) for motion detection. The major components for this sensor are an emitter and receiver assembly. They monitor frequency changes caused by moving objects (such as a person) through a phenomenon known as Doppler Effect. Unlike the PIR sensor, an ultrasonic sensor does not require a line-of-sight which makes it more suitable for occupancy monitoring in partitioned spaces. An ultrasonic sensor is more sensitive to minor motions, such as hand movements, when compared to a PIR sensor. Although, such improved sensitivity can lead to false switching (Floyd et al., 1995). This sensor is generally more expensive than a PIR sensor, with a single unit costing more than £13.

A typical microwave sensor has similar principle of operation as the ultrasonic sensor; the differentiating feature is the emitted signal, since it generates an electromagnetic signal at a frequency of 1-10GHz, and measures the frequency change of a reflected signal for occupancy detection (Runquist et al., 1996). The performance of this sensor is not affected by obstacles within its field-of-view in an observed space, although it can be prone to false switching, if adjacent spaces are occupied.

Video camera or networks of cameras are also used for building occupancy monitoring. Images can be observed by human operators or by the use of specialized computer software (Wusheng et al., 2008). While this method of sensing produces good accuracy, privacy may be a serious concern especially when occupancy counts are generated from images using human observers. However, analytical algorithms used for processing occupancy counts from images are still in the early stages of development (Benezeth et al., 2011).

Another method for building occupancy sensing is the use of biometric sensors which take measurements of human physiognomy for identification (Frischholz and Dieckmann, 2000, Dugelay et al., 2002), and produce an electrical signal. These sensors use algorithms to process images from occupants' physical characteristics, such as finger prints, eye-scan etc. These sensors contain an analogue to digital converter enabling it to digitise images, and store the digital information in memory so that it can verify the user next time he or she needs to authenticate their identity. Biometric sensors are mostly deployed for access control in buildings. High sensor cost is a major limitation for its widespread uptake.

Electromagnetic based sensors such as infrared sensors have potential use for occupancy sensing. The axiomatic people counter is a good example of this (Axiomatic-technology-limited). These are normally mounted at exit points to monitor occupant traffic. They use an infrared beam, with a transmitter and receiver pair mounted such that an infra-red beam is interrupted when occupants pass through the door. The sensor has good accuracy for establishing occupant presence but may be unsuitable for counting numbers of occupants, since the sensor is unable to detect multiple people crossing the infrared beam. Starting price for a beam counter is £276.00 for a basic model, up to £844.80 for one with Ethernet connectivity and automated reporting (Axiomatic-technology-limited). It is rarely used in building services control applications. However, it does have widespread usage in industries for machine operations safety.

An acoustic sensor such as a microphone can also be deployed for sensing occupants' activities (Fogarty and Hudson, 2006), and is simple and cheap.

However, noise from sources other than occupants can result in false triggering. Performance limitations associated with single occupancy sensing technologies may have prompted manufacturers to combine different technologies. Some commercially available products combine PIR and sound, or ultrasonic with microwave, thereby reducing the likelihood of false switching. Hybrid sensors tend to offer improved sensitivity, accuracy and flexibility, but often come at a higher cost. These sensors can be effective in partitioned offices, although their performance has not been documented (Maniccia and Wolsey, 1998).

2.2.2 Indoor air quality (IAQ) sensing

IAQ sensing is usually based on either carbon-dioxide (CO_2) or total volatile organic compound (TVOC) sensing, or a combination of both. Some sensing units monitor both with BEMS connectivity. Both are described here;

- **CO_2 monitoring**

The majority of CO_2 sensors used in building monitoring utilise non-dispersive infrared (NDIR) technology, which is based on the infrared broadband absorption characteristic of CO_2 gas. A light source transmits light (non-dispersive infrared) through a selective infrared filter into a measuring cell, where an interaction between CO_2 molecules and light occurs. This leads to a rise in temperature proportional to the CO_2 concentration which is commonly measured electronically using photo-acoustic or photometric methods (Won and Yang, 2005). In the photo-acoustic method, a small microphone is used to monitor vibration of CO_2 molecules caused by their interaction with light. CO_2 concentrations are then determined from vibration levels using on-board processing in the sensor. The photometric method measures the temperature of CO_2 molecules to determine concentration levels. Portable indoor units such as Telaire Ventostat T8200DB CO_2 sensor manufactured by GE sensing, which costs about £250/unit are widely utilised for ventilation control in commercial buildings (GE-Sensing). NDIR sensors can last up to 5 years, although they have the potential to drift significantly. Poor CO_2 sensor performance has been reported in some commercial buildings. In a recent study, the performance of 44 CO_2 sensors were investigated and results showed that measurements varied widely, sometimes in hundreds of parts per million, prompting recommendation from the author on the need to use more accurate CO_2 sensors, and calibration procedures (Fisk, 2008). Sensor accuracy can also be affected by vibrations and air pressure changes (Schell, 2001). CO_2 sensors are also not sensitive to other air contaminants, and this may be a major limitation for CO_2 based IAQ monitoring.

- **TVOC monitoring**

Metal-oxide and photoionization sensors are common technologies utilized for TVOC monitoring in many buildings. Metal-oxide sensors usually consist of one or more metal oxides from the transition metals such as tin oxide, aluminium oxide, etc. They typically come in two styles; bead-type (in which the metallic oxides are processed to form a bead-type sensor), and the thick/thin –chip type (in which the metallic oxides are vacuum deposited on to a silica chip) (Won and Yang, 2005). Gas molecules from TVOCs dissociate into charged ions when in contact with the sensing element of the sensor which results in electrons transfer. A pair of biased electrodes is imbedded into the metal oxide to measure its conductivity change, which is proportional to the gas concentration. This sensor has a very long life span typically up to 10 years, and can be applied for sensing a wide variety of gases. These sensors are susceptible to gas interference since they are inherently non-specific in application (Chou, 2000). A typical metal-oxide sensor such as the Aerasgard rlq-series air quality sensor could cost up to £130.00 (S+S).

Many TVOC sensor manufacturers employ a photoionization technique, where the main sensor component is a lamp filled with low- pressure inert gas which serves as an ultraviolet (UV) light source. Ionization occurs when gas molecules absorb energy from UV light, and produces current signals (with magnitude proportional to the gas concentration in the room) measured at the sensor electrodes. Lamps are usually specified in electron-volts: Several lamp specifications such as 10.6eV, 11.7eV, 9.5eV, 8.4eV, can be used as the UV light source for photoionization sensors. 10.6eV lamp is most widely used, having the longest life span of approximately 6000 hours. This sensor detects all gases with ionization potentials equal or less than the eV outputs of their lamps. For example, a 10.6eV lamp can only be used for detection of gases with ionization potential less than 10eV. They offer fast response, high accuracy and good sensitivity for monitoring a wide range of VOCs at low ppm (parts per million) concentration. A dirty lamp window can produce erroneous measurements, hence frequent cleaning is advised, although this is largely dependent on the level of contamination in the room. It was reported that high humidity decreases the sensor's response by about 30%, when compared to dry air, although this effect can be avoided by cleaning the sensor lamp (Chou, 2000). These sensors often require calibration and zero adjustments before carrying out measurements so as to compensate for background conditions (Chou, 2000). Photoionization sensors may be a very sensitive, reliable and durable technology. However, a major limitation is that it is considered expensive, with a typical unit such as Ion science TVOC sensor costing about £1900 (Ion-science).

2.2.3 Temperature sensors

Thermistors are semiconductors which exhibit changes in electrical resistance when exposed to temperature changes. The two basic types of thermistor are the negative temperature coefficient (NTC), and the positive temperature coefficient (PTC). The former is best suited for precision temperature measurement and the latter for switching applications (Jain, 1989). They are manufactured from oxides of transition metals such as manganese, cobalt, copper and nickel. Thermocouples are devices made from two dissimilar metals welded together such that a small open-circuit voltage (normally in millivolts) is produced through a phenomenon known as the Seebeck effect. This voltage magnitude depends on the material and the temperature difference between the junctions. Both technologies can be utilised for ambient temperature monitoring, although choice is dependent upon the desired accuracy, and temperature range of the measurement. Thermocouples can operate over a wide range of temperatures compared to thermistors. Generally speaking, thermocouple measurements will be more precise (Anon, 2005b), providing that the measurement-circuit temperature is accurately known. Conversely, accuracy of thermistors may be slightly greater, with slightly reduced precision. Figure (2.4) shows typical thermocouple and thermistor sensors.

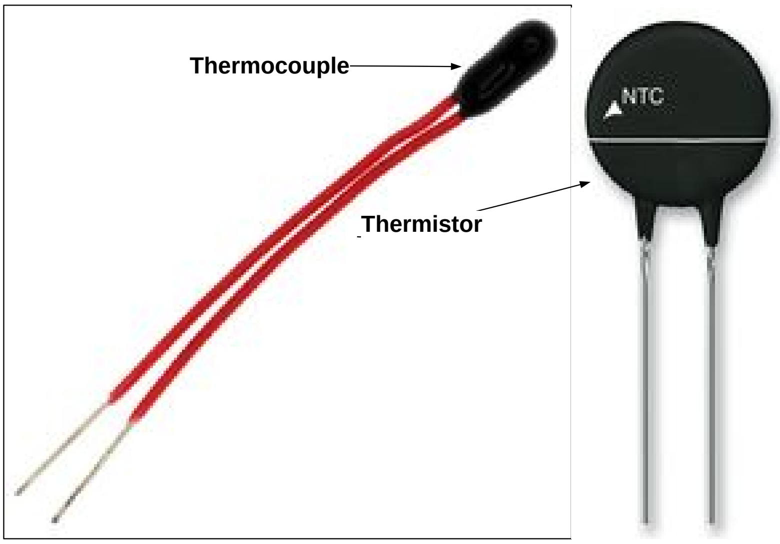


Figure (2.4): Typical thermocouple and thermistor sensors
Courtesy: Farnell (uk.farnell.com)

A Resistance Temperature Detector (RTD) operates through the principle of electrical resistance changes in pure metal elements (Hashemian, 2005). The element's resistance increases with temperature in a known and repeatable manner. Due to its low cost, and ability to measure point heat sources in a manner similar to a Thermistor, the RTD is slowly becoming more popular. Sensing circuitry is slightly harder to build due to a low sensitivity (though accuracy and precision are equivalent) compared with thermistors. This may also affect the use of the devices in areas with high levels of electromagnetic interference.

2.2.4 Relative humidity sensors

This section concentrates on the principal technologies for *automatic* monitoring of relative humidity (RH) in buildings. Reference devices, such as the gravity train hygrometer, or gravimetric hygrometer, are not considered here. These sensors are mostly used to control indoor RH levels especially in residential buildings, where they form an intricate part of extraction systems used to ventilate spaces when humidity levels are high (IEA, 1997). Semiconductor based RH sensors tend to show greater accuracy and reliability, and their prices are also affordable (Roveti, 2005). Capacitive RH sensors utilise thin film polymer or metal oxide that change capacitance with changes in humidity levels, and use an appropriate signal conditioning circuitry to measure humidity (Figure 2.5). The sensing surface is normally coated with a porous metallic oxide to avoid contamination or exposure to condensation. Self-logging version such as the HOBO relative humidity U-series loggers are readily available (Onset-Corperation).

Figure (2.5): Capacitive humidity sensing element
Courtesy: Farnell (uk.farnell.com)

Resistive humidity sensors measure the change in electrical impedance of a hygroscopic medium such as a conductive polymer, salt, or treated substrate. Resistive sensors are interchangeable (sensors are generally field replaceable), usable for remote locations, and cost-effective. Resistive RH sensors may suffer drift in condensing environment if water soluble coating is used for the sensing element, and does not show the same good long-term stability when compared to capacitive ones.

A thermal conductivity RH sensor comprises of two matched NTC thermistor elements in a bridge circuit. One of the thermistor is hermetically encapsulated in dry nitrogen while the other is exposed to the observed environment. Thermal conductivity (or absolute humidity) sensors measure absolute humidity levels by quantifying the difference between the thermal conductivity of dry air and that of air containing water vapour. The difference in heat dissipated between both thermistors results in a resistance change that is proportional to absolute humidity levels. For high temperature applications usually above 93^0C, thermal conductivity RH sensors are utilised, and would normally outperform both capacitive and resistive types (Roveti, 2005).

2.2.5 Power consumption

Typically, electricity meters are deployed for monitoring power consumption in buildings. An electricity energy meter usually measures voltage and current flowing through an electrical system in a non-invasive manner for power monitoring. Meter readings may be taken manually at periodic intervals, usually once a week or more typically once a month, doing this at a higher frequency is not practical. Large scale deployment of energy metering systems is seen as crucial for energy management in a non-domestic setting (O'Driscoll and O'Donnell, 2013), such a system often allow for continuous and automated monitoring on a more frequent basis. However, its implementation presents some challenges (such as the determination of number of installed meters, meter location, data interpretation technique and overall energy saving benefit), some which were addressed in O’Driscoll et al. (2012).

Many electricity meters produce pulse outputs which correspond to a certain amount of electricity passing through them. Pulse outputs may be dry contact in nature, meaning that the output is essentially a switch. Some manufacturers often state the nature of the pulse output: dry or wet contacts, as such information is necessary to determine the extra circuitry needed before the data logger. For instance, meters with mercury wetted contacts require some circuitry to electrically ‘de-bounce’ the signal, since more than one signal could otherwise be picked up, per pulse due to vibration of the switch contacts. A basic logic gate circuitry or a solid state output can be applied to address this. Specification and explanation of terms for pulse outputs are given in (BSI, 1999) and (BSI, 2002) respectively. Where a pulse output meter is fitted, it can be easy and straightforward to add a data logger with no further disturbance.

Current transformers (CTs) are also useful for monitoring building electricity usage. The most commonly utilised ones are the solid core and split core types (O'Driscoll and O'Donnell, 2013). Both are based on the principle of Faraday's law of induction to measure current flowing through an electrical circuit. The split core types tends to be less accurate than the solid core variations however, spilt core CT's can be installed more easily (Koon, 2002), and may be ideally suitable for remote monitoring applications without any disruption to a power system. CTs can be used in the case of either single or multiphase power consumption monitoring. The live or neutral cable (not both) usually goes through a CT's opening for both cases. Some CTs may be in the form of low-cost transducers, with their outputs often requiring signal conditioning, while others such as 'Clamp Meters' have inbuilt signal conditioning circuits, and are commonly applied for spot measurements rather than any long term monitoring. Multiphase units with pulse output which is compatible with conventional data loggers are also available. Several CT based 'plug in' power meters are available; a good example is the 'Watt Up' power meter (mostly marketed in US and Canada), which is capable of monitoring power factor. In Europe, smart 'plug in' units have also been introduced in to the market. For instance, Watteco units (with data logging capacity) (Bertrand, 2001), can distinguish between electrical loads plugged in, by detecting the unique electrical signature of each appliance, using proprietary signal processing techniques.

2.2.6 Luminance

Luminance data are necessary when carrying out robust building surveys (Mills, 1993), such data are useful for building performance and energy usage studies (Cohen et al., 2001), and for energy use prediction (Stokes et al., 2004). Previous research has proposed huge potential energy savings through widespread of use of day-lighting in buildings (Fontoynont et al., 1984), although, this is often neglected (Gakovic, 2000). Daylight-linked automatic lighting control is seen as a promising option for improved lighting energy management (Li and Lam, 2003). Walk through surveys (in places such as shopping malls) are particularly useful for identifying illumination types in use (Lam and Li, 2003), and can be less time consuming compared to practices which would involve querying electrical installation contractors, or individual shops. A more scientific approach may employ the use of handheld light meter which can monitor visible and UV light, as well as IR in the range of 350nm- ~40μm (Anon, 2005c). Figure (2.6) shows a typical handheld light meter with a digital readout.

Figure (2.6): A typical lightmeter
Courtesy: RS components (uk.rs-online.com)

2.2.7 Water flow

Proper understanding of water use patterns can provide compelling insights on space use (Beal et al. 2013), from which occupancy can be inferred. As with electricity and gas meters, water meters can be read manually with the same restrictions. Most water metering systems requires pluming work for set up, and typically produce outputs compatible with convectional data logging devices. For instance, a low-cost system marketed in the UK for £79 can easily be mounted on the pipework (Anon, 2005a), but would require brief disruption of water supply during installation, hence it is seen as minimally invasive .

Ultrasonic meters can be applied for flow measurements, where flow intrusion is not an option. Prices of these devices are gradually falling due to advancement in digital technologies, for example Texas Instruments 'industry standard' TMS320 series. Two basic formats of ultrasonic meters are available – Doppler (usually applied for contaminated liquids) and transit-time (for clean liquids). Clamp on units are readily available, and do not require cutting of pipes for installation. Figure (2.7) shows a typical ultrasonic clamp-on unit arrangement.

Figure (2.7): Ultrasound transducer clamping arrangement
Courtesy: Procon Systems Inc.(www.proconsystems.com)

Central heating water may be classed as a 'clean liquid'. Transit-time ultrasonic flow meters usually makes use of two transducers (forming a transmitter and a receiver pair), and measures the time it takes for an ultrasonic pulse transmitted from one transducer to go through a pipe's cross section, and be received by another transducer. This time of travel is proportional to flow rate (Figure (2.8) - Ultrasonic flowmeter). Deployment of these meters may be limited by pipe sizes (for example diameters of 6 – 20mm and 20 – 200mm requiring different sets of sensors) - requiring separate sensor heads, which often comes at a significant cost penalty. Besides, flowmeters would require calibration once new heads are fitted.

A detailed summary of the operational guidelines for ultrasonic flow metering can be found in Sanderson and Yeung (2002), while current state-of-the-art of this technology is given in Lynnworth and Liu (2006). Ultrasonic meters may offer two channels of flow metering and two channels for temperature metering, combined with a datalogger, to instrument, for example, send and return pipelines for a central heating boiler.

Figure (2.8) - Ultrasonic flowmeter
Courtesy: rshydro (www.rshydro.co.uk)

Doppler flowmeters are commonly applied for metering dirty or contaminated liquids (which require fluids with a minimum concentration of 100 ppm of solids or bubbles having minimum size of 100 microns), such as Agricultural water, Drilling mud, etc. They may also be useful for metering useful greywater or waste water output from buildings (for example from canteens), to assess opportunities for heat recovery, or to monitor water input for industrial uses (e.g. river water extraction).

2.2.8 Gas consumption

The vast majority of gas meters in domestic applications, including the UK are the diaphragm displacement meters (Cascetta and Vigo, 1994), although rotary displacement meters are also significantly in use. Both meters are purely mechanical, and are typically read manually at monthly intervals for billing purposes. Diaphragm meters measure gas flow directly from positive displacement (which drives a mechanical counting index indicative of volumetric flow) caused by filling and emptying of one or more measurement chambers. While, rotary displacement meter measures volumetric flow from gas displacement which occurs due to two impellers rotating at opposite direction to each other within the meter's internal housing. These devices may suffer high pressure losses (due to wear of their moving parts), and are unable to provide instantaneous flow-rate values (Hazlehurst, 2009).

With ultrasonic gas meters excessive pressure drops are rare and also offers non-invasive flow measurements using 'clamp-on' units, otherwise they are minimally invasive (Lynnworth and Liu, 2006). Ultrasonic gas meters usually consist of an emitter and a receiver pair which measures the transit-times (which is proportional to gas flow) of ultrasonic pulses between transmission and reception. Non-invasive ultrasonic flow metering is widespread, although metering can be affected by changes in pipeline wall roughness for fluids such as natural gas (Calogirou et al., 2001). Current state -of-the art smart meters have wireless automated meter reading (AMR), two-way communication and security capacity (Rouf et al., 2012). However, replacing existing mechanically based gas meters with ultrasonic types can be labour intensive, costly and cause disruption of gas service (Tewolde et al., 2013). A more convenient and less expensive approach is to non-invasively retrofit existing meters with modules to facilitate AMR capabilities. Several different AMR retrofit modules are already in use (Fischer, 2002), (Payne and Lien, 2011).

2.2.9 Airflow

Airflow dynamics within HVAC system can be crucial to performance of HVAC systems. Consequently, proper selection of airflow measurement devices becomes necessary. Traditionally, handheld vane anemometers (Various, 2005) and Pitot-tubes are widely utilised for airflow measurements (CIBSE, 1996). The working principle of a Pitot-tube is based on the principle of pressure differential of airflow measurements in building duct works. Vane anemometers (swinging vane being the simplest device) are placed in the air path within the duct work and airflow rates are measured from the vane rotation. These devices cannot measure air velocities below 1m/s without significant errors (Cheong, 2001). A Balometer can measure low volumetric flow rates with good accuracy (TSI), but will require the moving air to be channelled through a lightweight hood (capture hood). It uses a similar technology to an anemometer, and has been utilised to evaluate ventilation practices of occupants in a mechanical ventilated building (Park and Kim, 2012). Other instrumentation includes a thermal mass flowmeter, whereby the rate of heat removed by the flow stream passing a heated object is directly related to its mass flow.

2.2.10 Discussion of building instrumentation technologies

Although, the summary of technologies presented here is not an exhaustive list, hopefully it presents a comparative basis for existing building instrumentation technologies. This section is intended to give guidance on what sensing technology that is appropriate for a particular application. Sensor selection is largely driven by fitness to purpose, accuracy and cost considerations. It is clear from the researched literature, that most key aspects of building performance can be monitored and automatically logged. The use of state-of-the-art building instrumentations can facilitate robust characterisation of building performance, in terms of energy use and indoor comfort conditions. Although, some of the technologies platforms date back fifty years, advances in building instrumentation research continually challenge and improve the performance of sensors. With falling sensor costs, more BEMS are likely to make greater use of real time data for building controls. Many sensing technologies can be deployed as non-invasive monitoring platforms, such that more detailed and frequent monitoring of building plant and services are feasible, without undue disruptions to building operations. Various low-cost and non-invasive technologies for monitoring indoor environmental variables have been adopted in the multi-sensory data fusion approach used for occupancy estimation in this study.

2.3 Computational intelligence

Computational intelligence (CI) techniques are often referred to as soft computing. This section will focus on the commonly used computational intelligent techniques such as artificial neural network, fuzzy logic and genetic algorithms.

2.3.1 Artificial neural network (NN)

Artificial neural networks are biologically inspired models in which the output variables are computed from the input variables following a connectionist approach. These networks consist of a number of individual units called neurons. Connections between neurons have certain weights that are usually obtained using some learning rules. A typical NN has three interconnected layers: the input layer (where data are presented to the model), the hidden layer (where data processing is carried out) and the output layer (where results of a given input are produced). Each node has a predefined transfer or activation function. The output from each node is typically obtained by first taking the sum of the weighted inputs, together with a bias term, to form the intermediate quantity called the activation. In the second step, the activation is passed to the transfer function which returns the nodes final output. NNs have different structures; however, the multi-layered feed-forward architecture is commonly used, this thesis is limited to this type of NN. Figure (2.9) shows a typical structure of a feed-forward neural network.

NNs have the ability to perform a good amount generalization from the data on which they are trained. There are several NN training algorithms, examples including: Hebbian (Hebb, 1949), perceptron (Rosenblatt, 1958), and back-propagation (Rumelhart, 1986). The most successful and widely used is the back-propagation algorithm, where the network error is back propagated from the output to input layer. NN feed-forward model training process is such that input-output data sets are introduced to the network, only in a forward direction. Within the network, data are subjected to simple processing within its layers, and the weights of each neuron are adjusted in order to minimize the mean-squared error between the input and the target data, according to a specified accuracy index, or after the completion of a specified number of iterative learning processes. Once an NN model has been satisfactorily trained and tested, it can be used to predict output data from previously unseen input data.

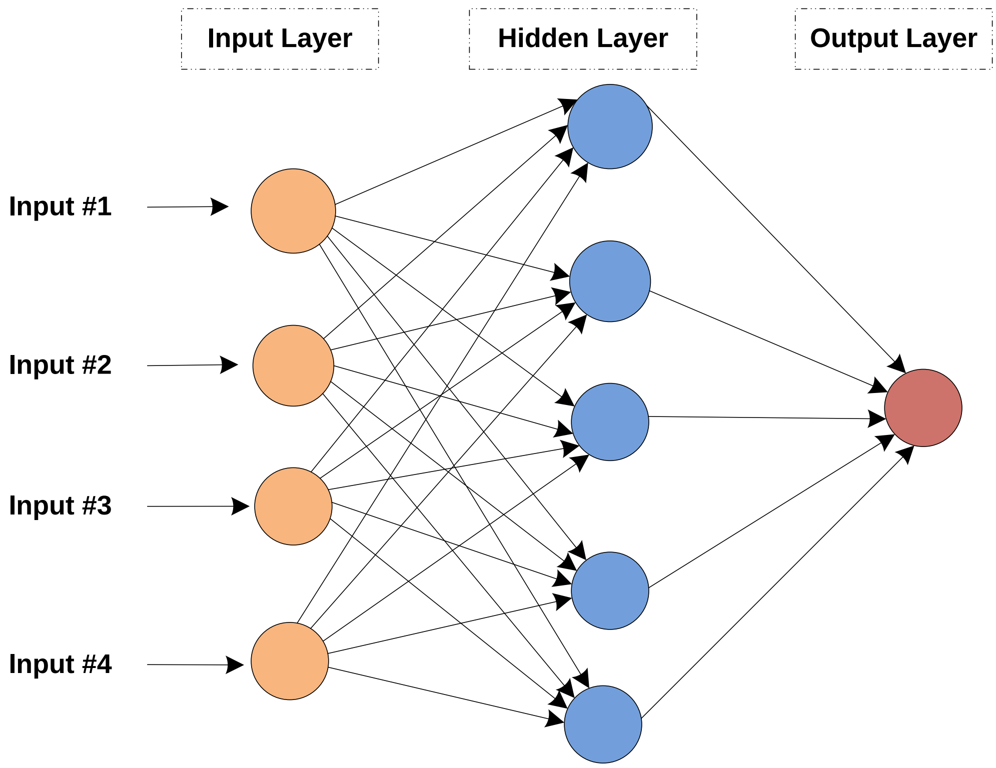


Figure (2.9): Typical structure of a neural network

NNs have been applied in buildings; a back-propagation model has been used for combining input features from various indoor environmental sensors (such as CO_2, PIR and acoustic sensors) for occupancy estimation in an open-plan office (Dong et al., 2010). Others have applied it for controlling HVAC heating coils by utilising a properly trained model to reset the PI control loop at a set point change (Delnero et al., 2001), and also for energy use prediction in retrofitted buildings (Li et al., 2009b), (Cohen and Krarti, 1995).

2.3.2 Fuzzy logic system (FLS)

Fuzzy logic was first introduced by Zadeh (1965). This approach mimics the imprecise reasoning and uncertain judgment of human beings. Unlike crisp set theory which uses the classical two-valued (zero or one) modelling of concepts, where sets have sharp boundaries and are mutually exclusive, fuzzy logic extends the classical logic theory by allowing intermediate truth values between zero (false) and one (true). Fuzzy logic allows partial belonging of any object to different subsets of the universal set instead of belonging to a single set completely. This partial belonging to a set can be described numerically by a membership function (MF) which assumes values between 0 and 1 inclusive.

A typical fuzzy logic system consists of three modules: fuzzification, inference engine and defuzzification. The fuzzification process involves mapping the input data to fuzzy sets defined by linguistic variables and membership functions. In the inference engine, the fuzzy sets are linked to one another using IF-THEN rules. In the defuzzification process, the fuzzy sets are mapped to output vectors. Figure (2.10) shows a typical fuzzy logic system.

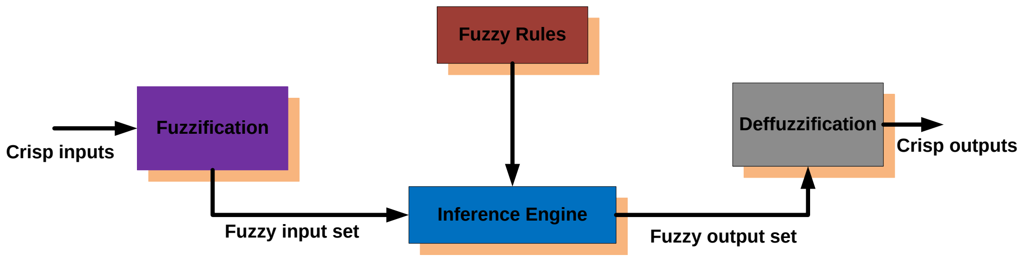


Figure (2.10): A Fuzzy logic system

A fuzzy logic model is trained using neural network, which results in a model known as the adaptive neural-fuzzy inference system (ANFIS). This was first introduced by Jang in 1993 (Jang, 1993). It combines neural network topology together with fuzzy logic (using a Sugeno model). ANFIS utilises the learning ability of neural network to define input-output relationships through hybrid learning to determine the optimal distribution of membership functions. These learning rules are based on a combination back propagation gradient descent error and a least square method. For example, for a first order –order Sugeno fuzzy inference system, the rules can be expressed as

Rule 1: IF x is A_1 *and y is* B_1 *THEN* $f_1 = p_1x + q_1y + r_1$

Rule 2: IF x is A_1 *and y is* B_2 *THEN* $f_2 = p_2x + q_2y + r_2$

Where x and y are the crisp inputs to the node i, A_i and B_i are the linguistic labels as low, medium, high, etc., which are characterized by convenient MFs and finally, p_i, $q_i$and $r_i$are the consequent parameters. $\overline{W_i}$ is the normalised firing strength in layer 3, and W_i is the firing strength in layer 2. Figure (2.11) shows a typical ANFIS architecture.

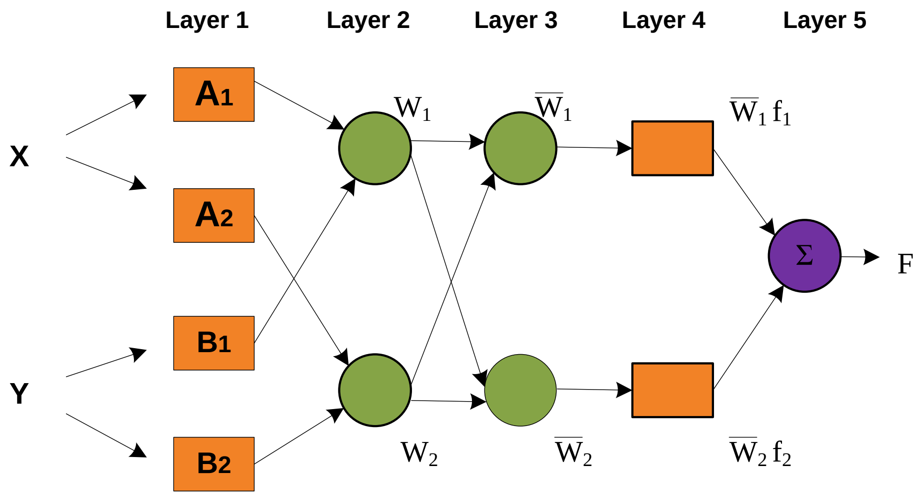


Figure (2.11): A typical ANFIS architecture

ANFIS is typically a five layered network with a similar principle of operation to a feed-forward network. In an ANFIS model, adaptive nodes are represented by squares while the fixed ones are denoted by circles. Fuzzification of inputs takes place in the first layer, and the resulting outputs are fuzzy membership grades. In the second layer, the nodes execute cross multiplication of all incoming signals from the previous layer, and the outputs represent the firing strength of a fuzzy rule. All nodes in the third layer normalises the firing strength of each rule. The output of each node in layer four computes the contribution of each rule to the model output. Layer five has just one fixed node which computes the summation of all outputs from layer four. The output of this layer denotes the overall output of the ANFIS model.

FLS’s have been applied in fault diagnosis for air-handling and conditioning units (Ngo, 1999), (Dexter, 2001), for indoor air quality sensing (Zampolli et al., 2004).

2.3.3 Genetic algorithm (GA)

GAs were originally developed by Holland (1975), but were made popular by Goldberg (1989). This algorithm uses the principles of natural selection to solve engineering problems. GA's are probabilistic in nature, and maintain a population of encoded solutions that develop, or evolve, towards a higher measure of quality, or fitness of these solutions. Figure (2.12) illustrates the operations of a typical GA.

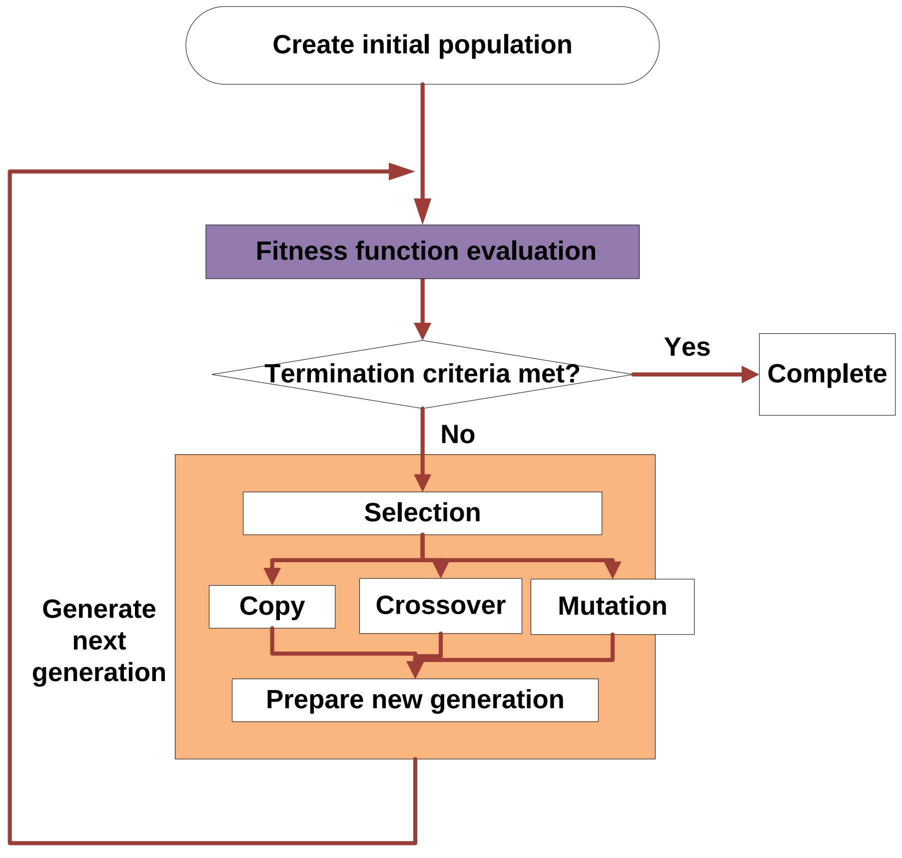


Figure (2.12): Typical structure of a GA

The GA search starts with an initial set of random potential solutions for an optimization problem. These potential solutions are often known as chromosomes (which are coded as binary or real strings); each is evaluated using some measure of fitness function which represents a measure of the success of the chromosome. A new set of chromosomes from the parent chromosomes are selected for breeding based on the fitness function which are then simulated using genetic operators (such as crossover and mutation) to produce a new generation. In the selection process, poor performing individual chromosomes are excluded, leaving fitter or better performing ones that have a greater chance of promoting the information they contain to the next generation. Crossover involves exchanging genetic information by swapping its bits within the parent chromosomes. This is a random process, where each gene of the chromosomes is selected from parent chromosomes to create new offspring. The mutation process adds new traits to the chromosomes, and helps a GA to avoid a local optimum solution. Once, an optimal or a near-optimal solution has been determined, a GA search can be terminated using a convergence threshold within a tolerable number of generations, processing time, or limitations in processing capacity.

Compared to neural networks and FLS, GA's have been sparingly used for building applications. However, GA's have been used to optimize energy usage in buildings (Lam, 1995).

2.3.4 Discussion of computational intelligence

CI techniques are inspired by nature and human reasoning and often are considered to have characteristics attributed to intelligence. These techniques are extremely useful in handling advanced non-linear practical problems. Data fusion techniques are drawn from various disciplines such as digital signal processing, statistical estimation, control theory, computational intelligence (CI) and classical methods**.** In this research, CI methods were studied and applied in the fields of multi-sensor data fusion and energy management.

Fuzzy logic system can handle uncertainties in sensor measurements. However, this was not been used in this work due to the huge time and computing requirements for training a FLS. NN was used for fusion of selected sensor data for occupancy number estimation, while GA was used to search the features space to obtain an optimal features subset for the fusion process.

2.4 Computational intelligence (CI) for building indoor environmental control and energy management

Building controls aims to reduce energy use and to maintain satisfactory indoor thermal and visual comfort, as well as indoor air quality. Standard control schemes, such as "on/off" and Proportional-Integral-Derivative (PID) are widespread in building applications (Loveday and Virk, 1992). For example, simple controllers such as thermostats have been used for indoor temperature regulation. These controllers could not prevent temperature overshoots, resulting in high energy use (Dounis and Caraiscos, 2009), and generally, do not provide optimal control. PID controllers were designed to address this problem. These controllers have a feedback mechanism, with constant parameters, and no direct knowledge of the observed environment. These controllers improved the situation, although, a wrong choice of gain can cause to the controlled system to be unstable (Dounis and Caraiscos, 2009). Generally, PIDs produce poor control performance for processes with large time – delay, significant noise and non-linearities (Li et al., 2006, Kaya et al., 2007). PID's performance can be improved by cascading multiple PID controllers (Kaya et al., 2007) or by combining feedback and feed-forward controllers (Thomas et al., 2005).

The use of advanced control schemes, such as computational intelligence (CI) based strategies (or in combination with PIDs), is considered as a promising approach to improve control system performance (Kukolj et al., 2001, Martins and Coelho, 2000). CI techniques are known to have coped well in noisy environments. They are adaptive in highly dynamic environmental circumstances, can be used to learn and generalize from examples, and can generate predictions at high speed (Hagras et al., 2008, Rafiq et al., 2001). Optimal (Zaheer-Uddin and Zheng, 2000), predictive (Chen, 2001, Henze et al., 1997) or adaptive (Curtis et al., 1996) controllers have been used to ensure satisfactory indoor thermal comfort, and also to limit set-point overshoots, thus reducing energy use. In order to use these controllers, a model (which is mostly non-linear in nature) of the building is necessary, and usually differs from one building to another. As a consequence, control schemes found in the literature always focus on a specific kind of building (Gouda et al., 2006). CI based control schemes have been largely limited to a research topic, as no massive industrial development has followed these scientific studies (Dounis and Caraiscos, 2009). The next section gives an overview of state-of-the-art applications of computational intelligence in building indoor environmental control and energy management.

- **State-of-the-art**

Computational intelligence is increasingly applied in solving engineering problems. Commonly used CI techniques are neural networks, fuzzy logic and genetic algorithm (Krarti, 2003). Others include multi-agents systems and ambient intelligence. They out-perform classical control systems (Kolokotsa, 2007). A brief overview of some CI contributions is discussed in the next section. However, a more detailed account of this can be found in (Dounis and Caraiscos, 2009), (Dounis, 2010), and (Kolokotsa, 2007).

2.4.1 Indoor thermal comfort control

It has been demonstrated in the literature that CI techniques have worked well for indoor climatic control. Thermal conditions inside buildings are highly non-linear; however satisfactory thermal comfort conditions for occupants have been achieved using CI techniques while optimizing energy use in buildings. For example, Jain and Ruxu (2005) designed a thermal controller that used a neural network and a thermal space model for a variable air volume (VAV) application to maintain the desired comfort level for space heating and cooling modes. Chu et al. (2005) developed a least enthalpy estimator (LEE) that combined thermal comfort level and theory of enthalpy in load forecasting for an HVAC system, in order to provide timely and suitable settings for a fan coil unit (FCU) fuzzy controller, which resulted in 35% reduction in the energy use. Calvino et al. (2004) used an adaptive-fuzzy network to improve the general characteristics of a classical PID temperature regulation system. They modified some control rules, aiming at determining a monotone "control surface" to guarantee better stability properties of the system. The addition of an adaptive network to the original model allowed for variation of parameters values regarding the integrative and derivative blocks. So doing, these parameters were dependent on the peak of the “step response”, which improved stability of the entire system.

For domestic heating control, Boait and Rylatt (2010) considered occupancy as a variable in their system design. Occupancy was inferred from electricity and hot water use. The system learnt occupancy pattern in the household and used this information to optimise energy use of the heating system. While this concept holds potential for improved energy efficiency in residential buildings, the technology is still in its early stages of development. It was unclear how this technology can be applied in non-domestic building where activities can be considered as highly dynamic in nature.

2.4.2 Indoor air quality (IAQ) control

Few studies in the literature have tackled IAQ control issues using CI-based controllers. Dounis et al. (1996a) investigated the performance of a fuzzy controller for IAQ control in naturally ventilated buildings. In their simulation, the controlled variable, CO_2 concentration, was maintained at satisfactory levels, while good stability of the control parameter (window opening area) was achieved. Dounis et al. (1996b) compared various schemes including on-off, PID, PI with dead-band and fuzzy control, for IAQ control in naturally ventilated buildings. Simulation results showed that the fuzzy controller reduced oscillations of the controlled variable, and generally provided better performance than others. Zampolli et al. (2004) developed a low cost miniaturized fuzzy logic based device for quantification of carbon monoxide (CO) and nitrogen dioxide (NO_2) in mixtures with relative humidity, and VOCs in a monitored environment. In this study, fuzzy logic system was used for pattern recognition to identify and discriminate concentrations as low as 20 parts per billion for NO_2 and 5 parts per billion for CO. There is still considerable scope for the application of CI techniques in addressing problems associated with IAQ in buildings.

2.4.3 Indoor visual comfort control

In visual comfort control, Yifei et al. (2009) proposed a neural-fuzzy smart system for the control of venetian blinds. The system was able to adjust the venetian blind position intelligently, and also automatically control indoor lighting level. Kurian et al. (2005) developed a computational model based on an adaptive neuro-fuzzy approach for predictive control of artificial lighting in accordance with the variation of daylight. Guillemin and Molteni (2002) developed an automatic and adaptive shading device controller based on a genetic algorithm capable of integrating user preferences with lighting control.

2.4.4 Indoor environmental control and energy management

Another dimension of intelligent control in buildings is an integrated approach of controlling the indoor environment as well as the energy use (Alcalá et al., 2005, Hagras et al., 2008, Dalamagkidis et al., 2007, Pargfrieder and Jorgl, 2002, Guillemin and Morel, 2001). Most notably among these works is that of Hagras et al. (2008), they developed an intelligent system for energy management in commercial buildings called intelligent control of energy (ICE). The system used a combination of neural network, fuzzy logic and genetic algorithm for optimization of energy use in an office building taking into account variables such as external weather conditions, internal occupancy requirements and building plant responses. ICE offers the possibility of choosing the most intelligent control technique for a particular set of environmental conditions. ICE can be integrated in to an existing BEMS to minimize energy use in real-time.

Advances in agent-based control systems and information technology have seen the birth of a new field called ambient intelligence which only came into existence in 2001 (Dounis and Caraiscos, 2009). It is a digital environment that is aware of its user presence and adapts to their needs depending on their behaviour (Rutishauser et al., 2005). While this technology is capable of delivering significant energy savings, it is still relatively new. The use of multi-agent based technology has also been explored in building energy management (Davidsson and Boman, 2005, Zeiler et al., 2006, Zhu et al., 2010, Sierra et al., 2006). An interesting use of agent based control for ambient intelligence was demonstrated in the work of Doctor et al. (2005), who proposed a life-long fuzzy learning and adaptation approach for intelligent agents that are embedded in an intelligent environment. The agents were deployed in an intelligent university dormitory (iDorm) to create an ambient intelligent environment. The agents were used to train a fuzzy logic controller that modelled the occupant’s preferences. In their work, learned users behaviour can be adapted in lifelong mode to satisfy different user and system objectives. In a five day experiment, learning of user preferences was adaptive since the embedded agents discreetly controlled the iDorm according to occupants’ preferences. The experiment was done in a non-invasive manner. Results indicated that the proposed approach performed better than other CI approaches while operating online in a life-long mode to achieve the vision of ambience intelligence.

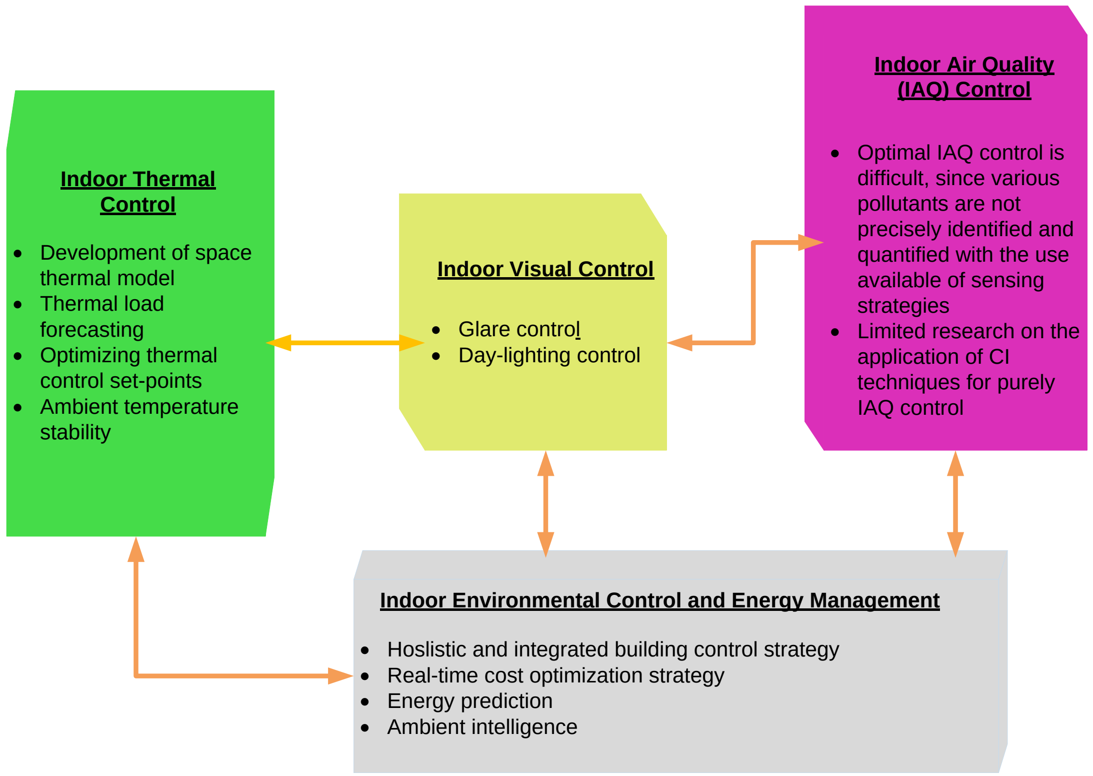


Figure (2.13): Some CI applications to building controls and energy management

2.4.5 Discussion of building controls

It is clear that there have been considerable advances in the application of CI techniques for indoor climatic control and building energy management. Some of the areas where CI is applied in building controls and energy management are shown in figure (2.13). Shortcomings of standard control schemes (e.g controlled parameter overshoot), may have facilitated the emergence of advanced control systems. A significant size of existing building controls still make use of PID controllers, coupling these with CI tools can help to improve their performance, and facilitate energy savings. In the reviewed literature, it was demonstrated that advanced control systems are adequate for maintaining acceptable indoor environmental comfort, as well as simultaneously ensuring considerable reduction in energy use. However, for majority of the studies, while consideration is given to the comfort requirements of occupants in the design of intelligent building control systems, the variable "occupancy" itself is rarely considered as an input parameter. In this research, a CI technique (neural network) is used as a predictive tool for occupancy numbers estimation, and the results being applied to drive a basic ventilation control strategy for an open-plan office space.

2.4.6 Challenges in building monitoring and control

The basic human need for indoor comfort has paved the way for the emergence of the building automation and control industry. However, there are some challenging aspects of building monitoring and control. These are discussed under the two headings: Sensor maintenance and ventilation control.

- **Sensor maintenance**

To maintain satisfactory indoor comfort conditions and energy efficiency, it is crucial for BEMS sensors to always maintain optimal functionality. However, this is hardly the case, proper building commissioning is sometimes not realistic, as the process may be shortened by delays in construction works, and perceived cost savings, and early occupation of buildings, hence proper testing of building controls are often considered as an afterthought (Levermore, 2000). Just one example of a typical error is when wiring errors are missed, resulting in a BEMS controlling one zone based on sensor information from another. Such mistakes often waste energy, decrease comfort, or both. BEMS sensors are usually checked during commissioning, although they often suffer significant drift over time, for instance, CO_2 sensors can drift up to 376ppm in one year (Shrestha and Maxwell, 2010). While this is a well-known maintenance issue, frequent recalibration of BEMS sensors usually does not happen (Apte, 2006). Continuous commissioning of BEMS sensors is vital to maintaining energy efficient building operations. Sensor maintenance is considered time consuming and costly, and is often rarely carried out. Painter et al. (2012) proposed the use of sensor overlay system to augment existing BEMS for proper commissioning. The use of virtual sensors for automated continuous commissioning is considered promising, far less expensive, and a more energy efficient approach (Li et al., 2011).Although, the technology is still in its early stage of development. Virtual sensors have been used for fault diagnosis and detection in buildings (Fan et al., 2010). A brief overview of the virtual sensing concept is provided in section 2.6.4.

- **Ventilation control**

Ventilation is a critical process for IAQ control, the supply of fresh outdoor air to remove air contaminants and odours is crucial in maintaining satisfactory IAQ inside buildings. Ventilation rates may be significantly reduced in buildings, so as to minimize energy use. However, in many cases, this may aggravate the problem of IAQ. Demand controlled ventilation (DCV) is a useful control strategy for reducing energy demands in buildings whilst also ensuring acceptable IAQ levels, ventilation rates are adjusted based on a measurable parameter (such as CO_2 concentration) that provides indicative information about the building IAQ or pollutant load at a given time. DCV is reported to have potential energy savings between 10% - 80% (Emmerich and Persily, 1997). Natural ventilation where applicable is far more energy efficient, although, comfort control in naturally ventilated buildings can be quite challenging, where air flows can be complex since individual zones (floors, rooms) are linked to allow air flow through the whole building (Linden, 1999).

ASHRAE Standard 62.1 establishes minimum ventilation rates for proper IAQ based on IAQ and contaminants and space occupancy (ASHRAE, 2004). Current ventilation systems utilize fixed occupancy schedules to control ventilation rates, which may not optimize energy use and comfort, depending on the occupancy dynamics of the space especially during unoccupied periods. DCV systems are increasingly being used for ventilation control in buildings, and most of them are CO_2 based. Occupancy numbers are inferred from CO_2 concentration in the observed space. Although, most current UK and European guidelines are based on maintaining specific CO_2 levels as per standard (EN15251:2007). Occupancy has a direct link with ventilation rate, although in many real buildings, occupancy and ventilation rates may not be stable for sufficient periods to enable an accurate estimation of ventilation rate from CO_2 data (Fisk, 2008). CO_2 based DCV can ignore other air contaminants including volatile organic compounds (VOCs) that may be present even when the CO_2 levels are low, since CO_2 sensors are insensitive to other air contaminants.

Considerable advances in building instrumentation have been made in recent years; such as the introduction of IAQ sensors (often known as VOC sensors), which monitor indoor pollutants such as offensive odours, smoke, and out-gassing from materials such as wall coverings, carpets, adhesives etc, which have impact on occupants' health and comfort. VOCs have been suggested to be a major cause of Sick Building Syndrome (Wang et al., 2008), such that monitoring should be expedient where this is suspected. Zitting (1998) provided a good summary of health effects on human beings of some common VOCs found in buildings. VOC sensors are increasingly receiving attention in DCV applications, due to its falling price compared to CO_2 sensors (which are relatively more expensive), offering 50% cost savings per sensor at installation (Painter et al., 2012).

Many instruments that have the ability to measure very low VOC concentrations, such as gas chromatography are considered large, bulky and costly (Pejcic et al., 2007, Wolkoff and Nielsen, 2001), which makes them unsuitable for use in building control applications. Existing VOC sensors are unable to distinguish between harmful air contaminants and benign chemicals. Attempts have been made to use

TVOC as a control parameter for DCV application. TVOC can be described as the sum of concentrations of all VOC components present in a sample which can be measured in real-time (Painter et al., 2012). A VOC sensor uses a tin metal oxide sensor for detection of TVOC or VOCs, and measurements are signal conditioned to produce a dimensionless output, which can be compatible with existing BEMS, although manufacturer's data on the relationship between sensors output voltage and VOC concentrations are not generally published (Agnello, 1999). It is difficult to calibrate these sensors to a known gas concentration as they produce different responses to different gases (Schell, 2008). This unselective response may have ramifications for applications in DCV. The sensitivity of an IAQ sensor is largely affected by temperature and relative humidity; such that seasonal variations affect longitudinal analysis (Schell, 2008).

Using VOC sensing for DCV application may be difficult to get right, particularly in new buildings, where initial VOC concentrations are likely to be very high especially in the first few months after construction, due to out-gassing from building materials such as furnishing, paints etc. The control sensor is normally tuned to provide adequate ventilation based on those high VOC levels, and may not be altered within the building life's cycle, even when VOC levels are lower (Painter et al., 2012), which may lead to occupants' dissatisfaction with IAQ. The case for TVOC as a control parameter in DCV applications has certainly not been advanced based on findings in the work of Potter and Booth (1994), who reported that VOC sensors used in their study were neither 'predictable nor reliable', i.e "no useful response was detected for occupancy or foodstuffs". The multiplier effect varies for different VOCs in the test area, but does not appear to be fully understood, and manufacturers' data is unclear (Painter et al., 2012). It is therefore difficult to specify a TVOC value that would ensure VOCs within a space are appropriate. Andersson et al. (1997) concluded that "their group cannot recommend the present use of TVOC as a risk indicator for health effects and discomfort problems in buildings". While there may be some sense in using TVOC as an indicator for space occupancy, this only tends to work in environments (such as kitchens, workshops, etc.) where VOC levels are strongly linked to occupant activities (Painter et al., 2012). In environments, such as new buildings, with the aforementioned out-gassing concern, TVOC for ventilation control is not suitable (Wolkoff and Nielsen, 2001). Further research is recommended in establishing VOCs for occupancy driven DCV (Won and Yang, 2005). A hybrid CO_2/VOC approach appears to offer a good compromise for controlling both occupant and building related contaminants in office buildings. In many applications, CO_2/VOC can complement each other, in ventilation control, such that this approach is considered promising. However, while the debate for appropriateness for VOC/CO_2 is still ongoing, each has its own peculiar application.

2.5 Occupancy driven building operations

The effectiveness of occupancy –driven HVAC operation and occupancy based switching of office electrical appliances; power management of office appliances are presented in the sections 2.5.1 and 2.5.2 respectively.

2.5.1 Effectiveness of occupancy –driven HVAC operation

Current HVAC systems in office buildings are usually operated based on fixed schedules, assuming maximum occupancy during occupied hours (typically between 9.00am and 6pm), and zero occupancy during nights and weekends. Clearly, this policy will not maximise energy savings, and does not consider periods when buildings are partially occupied. For instance, during the day, individual offices maybe in use regularly while other rooms such as conference rooms may be left empty or used semi-regularly.

With accurate real-time occupancy information, various demand-driven HVAC control strategies can be implemented. For instance, Agarwal et al. (2010) proposed that maintaining the ambient temperature other than what is specified by ASHRAE standards can ensure potential energy savings. In their case, the HVAC system was throttled back, maintaining a room temperature of 26.1^0C during unoccupied period, and 22.9^0C for occupied periods. Simulations yielded a 10 - 15% reduction of the daily HVAC energy use. In their recent work (Agarwal et al., 2011), the HVAC system was driven by real-time occupancy estimates from a wireless monitoring system, turning it off during unoccupied periods, and turning it on if the temperature goes over 24.4^{0}C, or under 18.9^{0}C during occupied periods. At the floor level, energy savings of up to 15.73% for HVAC electrical energy, and 12.85% for HVAC thermal energy was achieved. Pavlovas (2004) proposed a strategy to maintain lower ventilation rates during unoccupied periods, while keeping ventilation rate at maximum value when the building was occupied. Energy savings of up to 20% for ventilation was reported in the simulation. Yang et al. (2011) suggested an approach for supplying minimum airflow rates to each room as per ASHRAE standards based on occupancy information. This approach yielded up to 15% energy savings, when applied to an office space. Erickson et al. (2009) argued that outside air volume can be dynamically adjusted to match real-time occupancy loads in each room, instead of assumed fixed occupancy schedules and maximum design occupancy. HVAC energy savings of up to 14% was reported.

Once real-time occupancy changes are detected, associated changes in the heat loads can be calculated, such that HVAC systems respond immediately. Occupancy information can be used to control HVAC systems so that they respond to dynamic heat loads in a timely manner, before room temperature variations are detected by thermostats, thus ensuring improved energy savings. Tachwali et al. (2007) categorised cooling airflow rate into three levels- low, medium and high, so as to apply different set-points for each space. Based on real-time occupancy, the authors determined the cooling rate to be applied for each room. Energy savings up to 50% were achieved in the simulation.

HVAC systems can also be adjusted to suit individual comfort needs, if their localized room occupancy and preferences is known in advance. The work of Klein et al. (2012) supports this argument. A multi-agent based system which simulated the heating/cooling and ventilations of rooms based on detected occupants, and their preferences were proposed. During occupied periods, when there was nobody in the room, heating and cooling systems were turned off, while minimum ventilation was maintained. Otherwise, it was adjusted according to occupants' preference. This strategy generated up to 13.6% energy savings.

2.5.2 Power management of office appliances

The use of office equipment such as personal computers (PCs), printers, fax machine and servers constitute a significant part of the total energy consumed in commercial buildings. Electrical appliances use about 40% of office building electricity (Dounis and Caraiscos, 2009). Energy use of one PC can be relatively low, however, total energy use of PCs used in our homes and offices has highlighted PC usage as a target for energy conservation initiatives (OGC, 2008). Office equipment will be the fastest growing segment of commercial building energy use within the next decade (Webber, 2007). Presently, there are about 950 million desktop computers in the world (Somavat et al., 2010). The average PC power consumption is estimated as 60W (Walker, 2009), and with numbers in use expected to reach two billion by 2015 (Webber, 2007), the problem of energy conversation in commercial buildings is further compounded. Occupancy-based switching of electrical office equipment can contribute to efforts aimed at improving energy efficiency in office buildings.

Overall office equipment energy use can be better managed, if its usage pattern can be determined. PC's are sometimes left running by users when not in use, thereby wasting energy. Webber et al. (2006) reported that 52% of the equipment in their survey was left switched on when not in use. Power managing PC's is especially most valuable, in cases where users accidentally leave their computers on. A power management function in office equipment can be activated to reduce its energy use when not in an active state, although it is poorly implemented in commercial offices. For example, less than 10% of US-based PC's activate power management functions to take advantage of the energy saving potential (Walker, 2009), due to factors such as compatibility issues between different software components in PCs (Korn et al., 2004, Korn et al., 2006), general misconception, and lack of proper information on power management (Fujistu-Siemens, Raj et al., 2009). For example, in an inspection of 183,000 monitors worldwide carried out by Hewlett Packard, it was reported that one-third of monitors were not set to take advantage of their energy saving function due to fear of decreased performance (Hewlett-Parkard, 2006).

Computer manufacturers have different approaches power management, but the principle of operation is similar across all designs. The major difference however, is how the power management features in them are controlled. The ENERGY STAR office Equipment program which is close to an industry standard aims to address this by ensuring a sustained campaign for the provision of accurate information, and also support the implementation of power management in office equipment (Roberson et al., 2002). Power consumption levels are usually described using different terminologies between *on* and *off* level. For each successive low-power state, more hardware components of a computer is slowed down or turned off. The ENERGY STAR office Equipment program terminologies for the different power levels in office equipment are presented in table (2.3).

Table 2.3: Different power level in computers and monitors

	Desktop Computers	Monitors
Off	The Unit is plugged in (powered), the power button is in *off* position, and the power indicator is dark	
On	The power button is in the *on* position, the power indicator is green and the processor is idle	The power button is in the on position, the power indicator is green and the screen displays an 'empty' desktop; no application windows are open

Low-Power Levels		
Sleep	The lowest power level between *on* and *off*	The lowest low-power level between *on* and *deep sleep*
Deep Sleep	Not Applicable	The lowest low-power level between *on* and *off*

Source: (Roberson et al., 2002)

When the power management function is activated, the computer goes into a power saving mode after a period of inactivity. How often the computer enters a low power mode is dependent on the delay time chosen. Shortening the delay time for power managing office equipment was considered as an option to further increase energy savings, Kawamoto et al. (2004) reported energy savings as much as 3.5TWH per year for office equipment such as PCs, copiers and laser printers. This represented about 2% of commercial electricity consumption in Japan. Mungwititkul and Mohanty (1997) examined energy saving opportunities for office equipment used in commercial buildings in Thailand. Electrical power load patterns of equipment were monitored in order to determine how much time was spent in the different power modes. The authors reported up to 25% annual energy savings at no extra cost. Reuse of computers has also been suggested as a feasible approach to achieve energy efficiency, old computers can be used as servers in office environments (VitaminCM, 2008, Moshnyaga, 2008). In addition, reductions in harmonics associated with electrical power supply of office equipment can also ensure energy savings, and provide increased capacity for building power systems to serve other loads (Moreno-Munoz et al., 2009).

Lastly, standby power mode is an energy saving mode in electrical equipment, however, energy waste as a result of equipment in this mode is a source of concern in the quest for improved energy efficiency and sustainability. Raj et al. (2009) estimated standby power consumption for typical appliances, and reported that energy savings was achievable by avoiding standby power consumption. The authors suggested that standby power consumption can be reduced through behavioural and technical methods. It was advocated that better consumer awareness and education on standby power consumption would assist in addressing issues with end user behaviour in adopting energy efficient practises. From a technical aspect, redesigning appliance circuits can reduce standby loads by up to 90%.

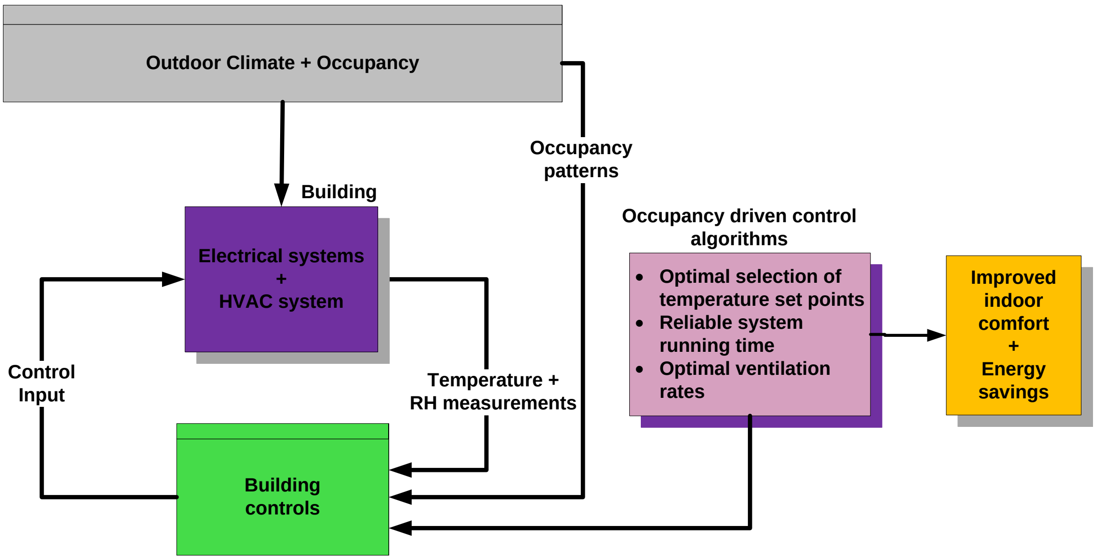


Figure (2.14): Occupancy driven controls

2.5.3 Discussion of occupancy driven operations

The biggest energy savings can be made in demand –driven HVAC systems, this is the main focus of the thesis. The link between lighting control and occupancy may eventually be explored in future work, although this has already received considerable attention. Figure (2.14) illustrates occupancy driven controls in a building.

Demand-driven HVAC control is an effective strategy to reduce building energy use. Several control strategies such as maintaining higher or lower temperatures depending on heat or air-conditioning in unoccupied periods, maintaining lower ventilation rates in unoccupied periods, supplying airflow based on occupancy, adjusting outside air volume based on occupancy, responding to dynamic heat loads on a timely manner and operating HVAC systems based on occupants preferences (Li et al., 2012), have been tested, and the results are promising. This list is not exhaustive, but highlights strategies mentioned in this work. Occupancy based switching is a useful strategy for reducing electrical equipment energy consumption in office buildings, although occupants rarely activate power management functions in appliances such as office desktop PCs. Various PC power management strategies such as shortening time delay for PC to enter power managed modes, reducing stand-by power consumption through improved education, and redesigning appliance circuitry were mentioned in this chapter. PC power management can be effective for reducing electrical energy use at unoccupied times, especially during after-work hours. The critical issue for implementation of these strategies is the availability of reliable real-time occupancy information.

2.6 Multi-Sensor data fusion and building instrumentation

Multi-sensor data fusion implies the combination of data from different sources (or sensors) to improve the reliability of a parameter measurement in an environment. Information from multiple sensors are combined to generate inferences that may not be possible from a single sensor alone (Hall and Llinas, 1997). Data fusion is hardly a new concept; it was first used in the US in the 1970s in robotics and defence. The Data Fusion Sub-Panel of the Joint Directors of Laboratories (JDL) was formed in 1985 to address some of the main issues in data fusion process in an effort to unify the terminology and procedure notably to improve communications between researchers and system developers (Hall and Llinas, 1997). Data fusion is currently applied in diverse areas such as military applications (Andersson and Ilestrand, 2007, Azzam et al., 2005), corrosion engineering (Jingwen et al., 2007, Zheng et al., 2007), food processing (al-Habaibeh, 2004, Huang et al., 2007), fault detection and diagnosis (Wu et al., 2010, Jaradat and Langari, 2009), automotive application (Herpel et al., 2008), welding (Cullen et al., 2008, Chen and Chen, 2010), medical sciences (Delis et al., 2009, Avor and Sarkodie-Gyan, 2009), and environmental applications (Zervas et al., 2010, See and Abrahart, 2001).

2.6.1 Fusion models

In this section, various fusion models which have been applied to aid the development of multi-sensor fusion systems are discussed with brevity. This is not an exhaustive list, but provides a representative overview;

- **The JDL data fusion model**

This model is widely accepted within the data fusion community. It was initiated by the US Department of Defence (DoD) to aid development of its military applications. It uses different sources of information ranging from sensor data to apriori information from databases to human input. The JDL model differentiates the data fusion process into five different levels (Hall and Llinas, 2001),see Figure (2.15). These are described more fully below.

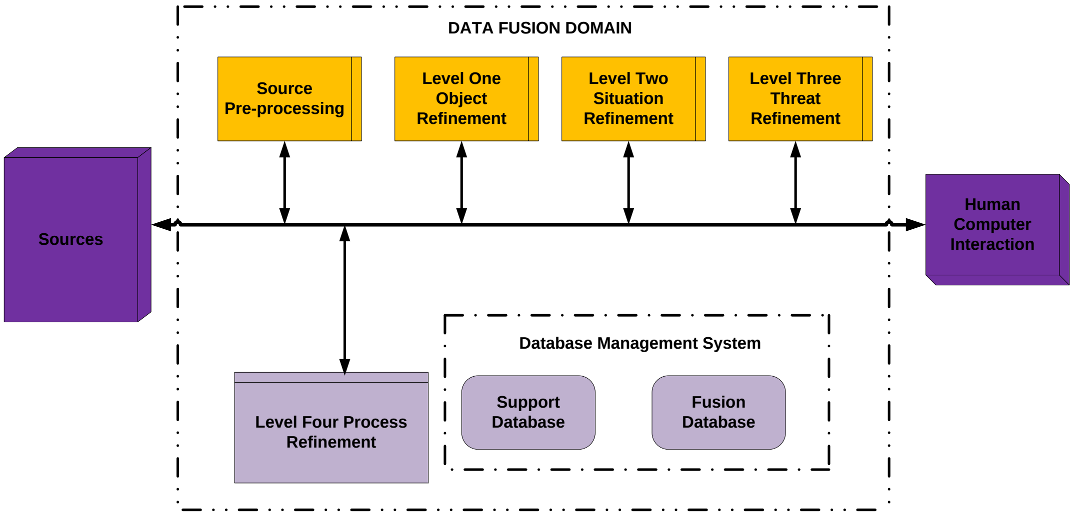


Figure (2.15): JDL model (Hall and Llinas, 2001)

Source pre-processing enables the fusion process to concentrate on data most pertinent to the present situation, thus reducing the processing load for a data fusion system by pre-screening and allocating data to the appropriate processes.

Level 1 (object refinement) is concerned with the combination of location, parametric, and identity information to obtain representatives of individual objects. This level performs the function of estimation and prediction of continuous or discrete states of objects.

Level 2 (situation refinement) attempts to examine the relationship between objects and observed events. This level provides an interpretation of the situation by aggregating objects into meta-objects.

Level 3 (threat refinement) projects the current situation into the future, to draw inferences from the current situation (level 2). This level also gives indication that a data fusion system may be operating in an adversarial domain (for example, an aircraft flying over hostile territory).

Level 4 (process refinement) concerns all the previous processes. It monitors performance of the fusion system, identifies potential sources of information enhancement, and optimises allocation of sensors in order to support the objectives of the mission.

The data management system performs data storage and retrieval functions to support the data fusion process, while the human–computer interaction provides a mechanism for human input and communication of data fusion results to the users.

The JDL model is intended to be general and useful for numerous application areas, although the model does not address multi-image fusion problems (Waltz, 1995) and does not support sensors involving multiple components (Hall and Ogrodnik, 1996).

- **The Dasarathy model**

Dasarathy proposed that there are three general levels of abstraction in fusion processing- the data level, the feature level and the decision level (Dasarathy, 1997).

Data is the sensor measurements from the environment which has not undergone processing (or maybe just undergone some pre-processing such as filtering).

Features are representation of information extracted from the data.

Decisions are inferences drawn based on the features.

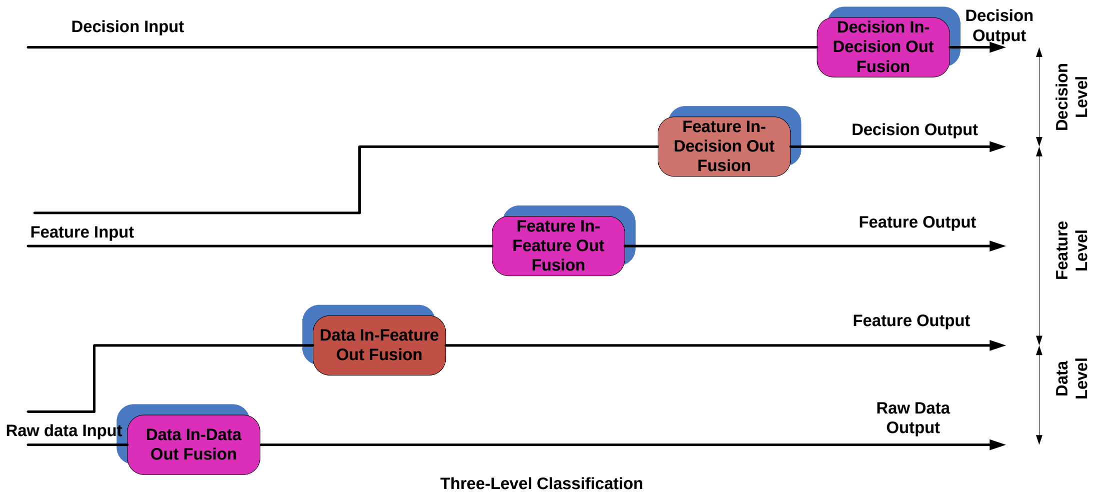


Figure (2.16): Dasarathy model (Dasarathy, 1997)

The three-level view was extended by Dasarathy into five fusion categories defined by their input/output characteristics as shown in Figure (2.16). The critical issue will be deciding whether to do fusion at the data, feature or decision level. The choice may be influenced by the goal of the fusion process and sensor arrangement. However, there are often compromises in performance whatever the method selected.

- **The Waterfall model**

The waterfall model was proposed by Bedworth (1994), and places emphasis on processing functions at the lower levels. Figure (2.17) shows the various processing stages in the model. Similarities can be drawn between this model and the JDL model, such that sensing and signal processing correspond to source pre-processing (level 0), feature extraction and pattern processing match object refinement (level 1), situation assessment is similar to situation refinement (level 2), and decision making corresponds to threat refinement (level 3).

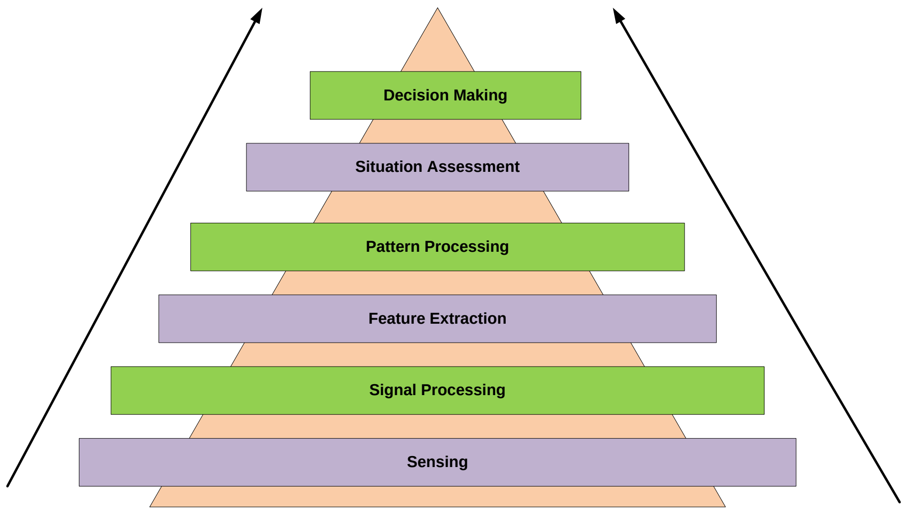


Figure (2.17) :Waterfall model (Bedworth, 1994)

The Waterfall model provides more precise analysis of the fusion process than other models. However, omission of feedback data flow constitutes a major drawback. It has been largely utilized within the defence community in the United Kingdom but has not been significantly accepted in other parts of the world (Bedworth and O'Brien, 1999).

- **The Boyd model**

The Boyd model was proposed by (Boyd, 1987) and has ever since remained an important concept in the military. The model contains a cycle containing four stages as shown in Figure (2.18). Bedworth and O’Brien (1999) made comparison between the Boyd and JDL model.

Observe: This stage is broadly comparable to source pre-processing (Level 0) in the JDL model.
Orientate: This stage encompasses functions of the levels 1, 2, and 3 of the JDL model.
Decide: This stage is comparable to level 4 (Process refinement) of the JDL model.
Act: Since JDL model does not close the loop by taking the actuating part of the interaction into account, this stage has no direct counterpart in the JDL model.

This model depicts the stages of a closed control system, and presents an overview on the overall task of a system. However, its structure is not suitable for the identification and separation of different tasks in a fusion process.

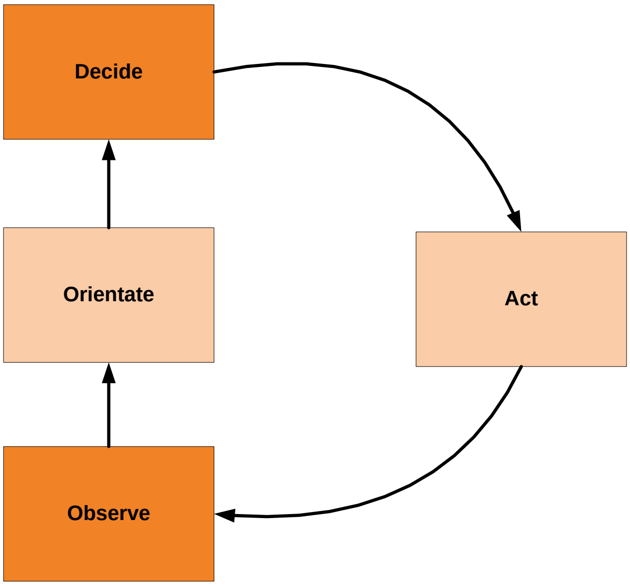


Figure (2.18): Boyd Model (Boyd, 1987)

- **The Omnibus model**

This model was presented by (Bedworth and O'Brien, 1999), it is more of an hybrid of Boyd, Waterfall, and Dasarathy models. Figure (2.19) depicts the general layout of the Omnibus model. Unlike in the JDL model, feedback is explicit in this model. It shows a cyclic structure comparable to Boyd's model, but provides a much more fine-grained structuring of the processing levels. The terminology used in this model makes it suitable for general applications and not limited to defence-oriented services. The model is intended to be used more than once in the same application at two different levels of abstraction. Firstly, the model is used to characterise and structure the overall system. Secondly, the same structures are used to model the single subtasks of the system.

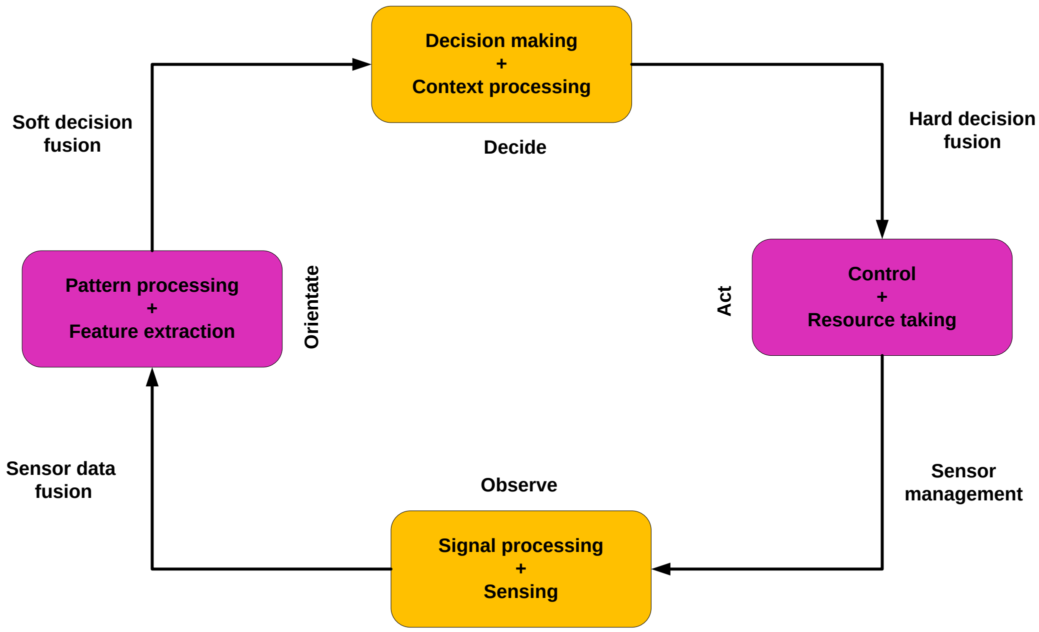


Figure (2.19):Omnibus Model (Bedworth and O'Brien, 1999)

In this model, the hierarchical separation of the sensor fusion tasks is sophisticated. However, it does not support a horizontal partitioning into tasks that depict distributed sensing and data processing. Therefore, this model does not sustain the break-down of tasks in to units that can be separately implemented, tested or reused for different applications.

2.6.2 Feature selection

With the rapid advancement of computer and database technologies, datasets are increasingly becoming larger in the number of input variables, instances and features. Many machine learning algorithms were not originally designed to cope with large amounts of irrelevant and redundant features (Guyon and Elisseeff, 2003). In the presence of these, learning models tend to over-fit data and their results become less comprehensible. Hence, combining machine learning algorithms with feature selection techniques to improve performance has become a necessity for many applications (Guyon and Elisseeff, 2003).

Feature selection is frequently used as a data processing step in the development of sensor fusion systems and data mining applications. Feature selection is a process of choosing a subset of features according to a certain evaluation criteria, such that the feature space is optimally reduced. It has become the focus of many research fields in recent years, and has been proven to be an effective step in removing irrelevant and redundant features, reducing data processing time, increasing efficiency in learning tasks, improving learning performance such as predictive accuracy, and enhancing comprehensibility of learned results (Blum and Langley, 1997),(John et al., 1994). Feature selection has been successfully applied in military (Andersson and Ilestrand, 2007, Azzam et al., 2005), medical sciences (Delis et al., 2009, Avor and Sarkodie-Gyan, 2009), occupancy detection (Dong et al., 2010), (Hailemariam et al., 2011) applications, to mention a few.

Generally, a large amount of building indoor environmental data is necessary to be collected in advance for development of an occupancy detection system based on the sensor fusion approach. Data can include temperature, lighting levels, CO_2 levels, sound, motion and relative humidity measurements. The presence of irrelevant and redundant features in the data can affect detection performance of the system. Therefore, feature selection becomes vital.

Feature selection algorithms can be broadly classified into two categories: The filter model and the wrapper model (Das, 2001),(John et al., 1994). The filter model relies on general characteristics of the training data to select features without involving any learning algorithm. Using the filter model, there is the risk of selecting features subsets which may not match the chosen learning algorithm. However, they are computationally cheap, and do not inherit any bias of a learning algorithm (Senthamarai Kannan, 2010). On the other hand, the wrapper model employs a predetermined learning algorithm to search the feature space, and determine features subsets (with the highest quality) based on its performance. The Wrapper model tends to provide superior performance than that of a filter based approach, as it finds features that are better suited for the predetermined learning algorithm. But it tends to be more computationally expensive and time consuming than the filter model (Blum and Langley, 1997). Thus, it becomes unpractical to apply a wrapper model for feature selection when the data set is large containing numerous features and instances (Hall, 1999).

Quite recently, researchers are combining the advantages of both approaches to form a hybrid or embedded model (Das, 2001). In this approach, the search for the optimal subset of features is built into the classification algorithm, and can been seen as a search in the combined space of feature subsets and the hypothesis. Just like the wrapper model, it tends to be biased to a predetermined learning algorithm, although it is less computationally intensive than the wrapper model.

2.6.3 Feature selection algorithms

Here, some representative feature selection algorithms are discussed with brevity. Filter models can be further divided into groups namely: The features weighting, and subset search algorithms, based on whether the goodness of features is evaluated individually or through feature subsets. Features weighting algorithms usually assign weights to features individually and rank them based on their relevance to the class, a good example of this is the Relief algorithm (Kira and Rendell, 1992). The key idea of Relief is to estimate the relevance of features according to how well their values distinguish between the instances of the same and different classes that are near each other, although this algorithm cannot handle redundancy in the features space. Many other algorithms such as Laplacian Score (He et al., 2005), information gain (Cover and Thomas, 1991) in this group face the same problem. The use of principal component analysis (PCA) analysis for feature selection has been long demonstrated in the machine learning literature. PCA involves the transformation of a set of features which are correlated in to a new set (principal components) which are uncorrelated. The first principle component usually contains as much variance as possible compared to the preceding principal components. All the principal components may not possess equal predictive strength. PCA can be useful for determination of relevant features for the development of occupancy detection systems (Yang et. al, 2012). However, pure relevance based feature weighting algorithms may not be adequate for optimal feature selection due to the issue of redundancy (Hall, 1999), since they are incapable of removing redundant features since redundant features are likely to have similar rankings or predictive power (Yu et al., 2004).

On the other hand, subset search algorithms address the issue of redundancy by searching for candidate feature subsets using an evaluation measure, which captures the goodness of each subset, and stops the process when an optimal (or near optimal) subset is selected (Liu, 1998). Consistency measure (Dash et al., 2000) and correlation measure (Hall, 1999) are known evaluation measures, which are effective for removing irrelevant and redundant features. Consistency measure attempts to find a minimum number of features that separate classes as consistently as the full set of features can (Dash et al., 2000). An inconsistency is defined as two instances having the same feature values but different class labels. Consistency driven feature selection process may be intractable if many features are needed to attain consistency, and may generate a strong bias towards consistency which can lead to over-fitting of training data (Hall, 1999). Correlation measure is applied to evaluate the goodness of feature subsets based on the hypothesis that a good feature subset is one that contains features highly correlated to the class, yet uncorrelated to each other (Hall, 1999). Correlation measure is by far more popular than consistency measure, although both can be used alongside different search strategies, such as exhaustive, heuristic, random etc, combined with the evaluation criteria to form different algorithms.

In this research, the filter model for feature selection is chosen over the wrapper model mainly because it is faster, and requires less computational resources. The correlation based feature selection proposed by Hall (1999), and implemented in Waikato Environment for Knowledge Analysis (*WEKA*) (Hall et al., 2009), is adopted for this research. To the best of the researcher's knowledge, application of this feature selection methodology to indoor environmental data for occupancy detection is new.

2.6.4 Virtual sensors for buildings

Individual building services components are increasingly becoming more efficient, for instance the rated efficiency of new residential cooling equipment has nearly tripled in the last decade (C.E.C, 2008). However, the operational energy can be degraded by 20% to 30% due to improper installation/commissioning and inadequate maintenance/repair (C.E.C, 2008). Problems that develop during operation are often ignored as long as comfort is satisfied, leading to inefficient operation (Cisco, 2005). Some studies have reported energy wastage between 15% and 50% due to faults or non-optimal operations (Katipamula and Brambley, 2005). One approach to improve building operations may be through the use of more reliable sensing, which could mean deploying larger numbers (possibly thousands) of sensors. Although, this would result in higher installation and maintenance cost, since building systems can be very large and complex, serving hundreds of zones with individual controllers to adequately characterize and monitor performance. A building with many indoor sensors may place heavy maintenance burden on its facilities department to ensure optimal functionality of each sensor. Without an effective quality assurance program in place, accuracy and reliability problems with sensors can quickly diminish the effectiveness of any existing BEMS. Virtual sensors produced from a fusion process with the potential for providing high-level performance monitoring information can be a more reliable and robust alternative for monitoring (Dodier et al., 2006), (Kusiak et al., 2010). These sensors may also limit the number of sensors needed for effective monitoring and control, thus ensuring overall sensor cost reduction, and minimal maintenance burden.

Virtual sensors sometimes known as “soft sensors” include any indirect method of determining a measureable quantity that utilizes outputs from other physical and/or virtual sensors along with process models and/or property relations (Li et al., 2011). High-value sensor information from virtual sensors could enable optimization and improvements of building operation not previously possible (Li et al., 2011). Although, some might argue that physical sensors can provide sufficient information for building controls, cost reduction in building monitoring may be a key driver motivating interest in this approach. A virtual sensor becomes handy, where it is difficult to measure a quantity such as in retrofit application (e.g. measurement of refrigerant flow rate or pressure), it can be more easily added to an existing system as opposed to a physical sensor (Li and Braun, 2007). For a variable such as building occupancy number, which can be difficult to measure, the concept of virtual sensing can be useful for developing more reliable occupancy sensors. For instance, the proposed occupancy sensor in this research utilise inputs from various physical indoor environmental sensors and a data processing model for occupancy numbers estimation.

Implementation of a virtual sensing platform often involves the following process: Data collection and pre-processing, model selection and the sensor implementation, as shown in figure (2.20). Proper data collection and pre-processing are crucial to the performance of virtual sensors. Data used to train virtual sensor models are collected from physical sensors, and may contain significant amount of noise. Different pre-processing methods are usually applied to the data for noise removal, so as to improve sensor performance. Modelling and training are the most difficult process in the development of a virtual sensor. With no shortage of models, selection is often more like a trial and error process, and often considered an “art”. However, consideration is given to the model accuracy and parameter estimation. Sensor implementation stage involves validation of model performance and incorporation in to an existing control system or application as a stand-alone device.

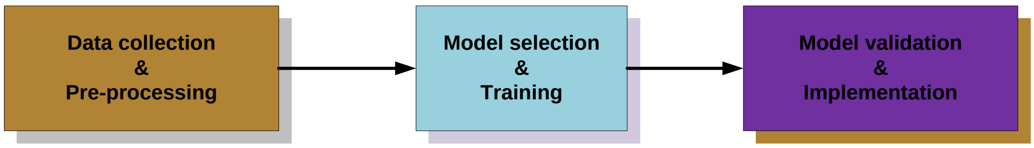


Figure (2.20): General steps in developing virtual sensors

Virtual sensors have been used in fault diagnosis and detection (FDD) (Li, 2009), parameter monitoring (Li and Braun, 2007), occupancy detection (Dong et al., 2010), (Meyn et al., 2009) etc. While virtual sensors produced from fusion systems have the capacity to provide more robust sensors, their performance could also be flawed if training data used for system modelling is corrupted (Hall and Llinas, 2001).

2.6.5 Discussion of multi-sensor fusion and building instrumentation

The concept of multi-sensor data fusion has been introduced. The fusion model in closest spirit to the concept of occupancy detection from indoor environmental data is that proposed by Dasarathy (1997), others have mostly been deployed in defence applications. Feature selection is an important data processing step in developing systems using machine learning algorithms. Irrelevant and redundant features in data may degrade the performance of an occupancy detection system. A brief overview of some commonly used feature selection algorithms has been presented, and the correlation based filter model has been identified as a particularly useful technique for feature selection. This choice is informed by its relatively low computational and processing time requirements compared to other feature selection models. In this research, an improvement of information gain, such that it is capable of handling feature –feature redundancy was implemented.

Sensor fusion has the potential to change the way building monitoring is carried out, such that more reliable real-time building performance data can be generated. With the growing need to reduce cost of building operations, whilst maintaining a comfortable indoor environment, the use of virtual sensing techniques can contribute to efforts aimed at developing automated continuous commissioning systems, such that sensor failures are identified real-time, and thus dramatically reduce the frequency at which specialists/technical staffs are called out to trouble shoot building energy systems.

2.7 Chapter summary

This chapter has presented state-of –the-art in building indoor environmental monitoring and control, focussing on occupancy detection. Through this review several points have been observed:

- The need for a heterogeneous multi-sensory occupancy detection system has been clearly advocated in the reviewed literature. A number of the occupancy detection systems examined have certain short-comings with respect to accuracy, cost, intrusiveness, and privacy. Sensor selection for occupancy detection has been informed by studies in the literature; with low-cost off –the- shelf sensor technologies being the preferred ones.
- In recent times, more and more methods have been applied to building occupancy detection based on both statistical and machine learning algorithms, such as HMM, NN, SVM etc. Many machine learning algorithms can be used for occupancy estimation, although model accuracy is largely dependent on training data. A neural network based fusion approach has been

adopted in this research for its fast computational time and simplicity to use. In order to examine, the robustness of the model estimations, it will be compared with other machine learning algorithms in section 5.9.

- Most multisensory occupancy detection systems in the literature have been built around the Dasarathy fusion model, where features providing occupancy information are extracted from sensor data before the actual fusion process. The features used in most occupancy detection systems are often peculiar for the observed indoor environment, and the sensor types deployed. Some of the features used have been designed along the lines of temporal analysis. In this research, the Dasarathy model has been adopted for its suitability to data collected in this research, and new sets of features were investigated.
- Correlation measure based analysis for feature selection has been identified as a useful method. A modified version of information gain theory, called the symmetrical uncertainty (SU) evaluation is selected for this study, as it addresses the drawback of information gain, which is known to be biased to data sets with more values. SU will be discussed in more details in chapter four.
- Clearly, convectional building control systems rarely make use of "occupancy variable" as input to control systems. The effectiveness of occupancy driven control of HVAC systems have been thoroughly researched. The importance of CI techniques for building controls has also been studied. These techniques are useful for developing multisensory occupancy detection systems. For the purpose of continuous commissioning, virtual building instrumentation (such as the occupancy sensor developed in this research), can contribute to efforts aimed at reducing energy use in buildings. However, further development of virtual sensing systems in buildings is recommended.
- The review also provided a fundamental understanding of different sensing technologies, highlighting their advantages and exposing their limitations. It also gives an insight into the cost-benefit consideration for sensor selection used in the research.

- From the review, it is clear that there is shortage of a well-defined method for developing occupancy detection systems. This research investigates a systematic methodology for determining the relevant indoor environmental variables necessary for development of a robust system for occupancy detection.

In conclusion, a heterogeneous multisensory approach for occupancy detection seems to have the best potential in addressing the short-comings of existing occupancy detection systems. This research aims to develop a detection system to provide reliable occupancy information.

CHAPTER 3

ADVANCED MULTI-SENSORY INSTRUMENTATION STRATEGY: EXPERIMENTAL SET-UP AND DATA ACQUISITION

3.0 Introduction

This chapter presents a detailed description of an advanced multisensory building occupancy instrumentation strategy. The strategy here is to overcome the limitations of systems previously described in chapter two, by fusing information from a network of low-cost and non-intrusive sensors for building occupancy estimation. This chapter is organised as follows: section 3.1 presents the research hypotheses, 3.2 describes components of the experimental design for occupancy estimation. Section 3.3 presents findings from a pilot experiment. Section 3.4 describes electromagnetic interference (EMI) mitigation strategy for CO_2 sensors deployed in the research. Section 3.5 presents the design and implementation of a custom sound sensor, while section 3.6 describes the process of occupancy validation. Lastly, section 3.7 summaries the chapter.

3.1 Research hypothesis

According to the literature review on building occupancy detection systems, interactions of occupants with their indoor environment clearly affect the surrounding climatic conditions through the emission of CO_2, sound, heat, moisture, and propagation of activities such as appliance usage (Dong et al., 2010), (Meyn et al., 2009), (Hailemariam et al., 2011), (Brown et al., 2011), (Cleveland and Schuh, 2010), (Page et al., 2008). These variables can be useful occupancy proxies and can complement each other for estimation of occupancy level in an indoor space. Hence, the following conjecture is formulated:

The combination of information derived from low-cost and non-intrusive indoor environmental sensors using machine learning techniques can provide reliable occupancy estimations in a naturally ventilated open-plan building.

3.2 Experimental design

The research question driving the thesis is the desire to explore which indoor environmental variables are relevant for development of a robust system for occupancy detection, with a view to reduce energy use in a naturally ventilated building using a sensor fusion approach. To answer this, it was necessary to design an experiment that can generate a representative data to carry out a robust investigation, in order to test and validate the analysis. The research design employed in this work allows for both gathering of occupancy and indoor environmental data, and testing the hypothesis in an open plan office setting, where a variety of environmental climatic sensors have been deployed and the data obtained reflect the phenomenon to be studied in its naturalistic setting. The instrumentation strategy implemented in this research uses low cost non- intrusive sensors.

The research design allows for a holistic investigation of the suitability of different indoor climatic variables for occupancy estimation using computational intelligence (CI) techniques. The rich environmental data set ensured by the research design support the exploratory nature of the research question. As established earlier in this thesis (chapter two), reliable building occupancy monitoring is difficult with existing technologies, and the experimental design used in this research is a direct attempt to address this gap. The outcome of the thesis therefore, is a new data processing methodology for building occupancy estimation, validated and tested in a real environment. The use of office equipment case temperature monitoring (which has not previously being examined for occupancy numbers detection) is also introduced.

Designing an experiment for building occupancy monitoring using indoor environmental sensors requires a carefully thought out and systematic process sometimes along conflicting lines. This design addresses issues such as monitoring locations to be used in the study, duration of the monitoring campaign, and factors that interfere with the indoor climatic data. The aim of the experiment is to generate reliable fine-grained indoor climatic measurements using low cost non-invasive sensors for the purpose of occupancy estimation.

- **Monitoring objectives**

1. To gather indoor climatic data for occupancy numbers estimation in various locations across a naturally ventilated university building.
2. To explore the relationship between different building variables.

3.2.1 Deployment environment

Several test areas were chosen for data gathering and system development within the Queen's building - this is an advanced naturally ventilated building which forms part of the De Montfort University campus in the English Midlands, and was constructed in the early 1990s. It was lauded at the time for being the first in a new generation of low energy buildings. It houses an engineering department, the Institute of Energy and Sustainable Development (IESD), laboratories, lecture theatres, and offices for research and administration. After building commissioning, energy use of the building was considered low at the time, although after several years of post-construction occupancy, it increased. This, in part can be attributed to an increase in the number of IT equipment in the building, and the presence of a radio station transmitter in the building. All the test spaces have been chosen because of their multi-occupancy nature, and also their indoor environmental conditions can be considered as heavily dynamic. In addition, they provide a good representation of various components, and activities within a typical university office building. All sensors were placed 1.5m from the ground as per standard practice (CIBSE, 2009). Placement of sensors was largely influenced by the physical configuration of the test space. Detailed descriptions of selected test areas in the building used for data collection are presented below;

- **Test area one**

This test area is an office kitchen used by around 39 people. The kitchen is fitted with a refrigerator, a microwave, an electric kettle and cupboards for storing beverages. Heating is supplied by four convective radiators. The area has two sets of windows, one facing the sun directly while the other is shaded by an adjacent building. The occupancy schedule in the kitchen is typical of an office setting, with light morning traffic, heavier in the afternoon, and peaking at lunchtime, and also light traffic in the evening. This is a fairly small space, and IAQ can quickly deteriorate at peak occupancy. Figure (3.1) gives a clear picture of this test area.

Figure (3.1): Test area one

In test area one, a HOBO UA-001-08 temperature sensor was fitted to the microwave enclosure. A PIR sensor provided motion detection, located in a corner, pointing towards the entrance. Sensors were also fitted to measure relative humidity and air temperature. All sensors were still under the calibration warranty period, and were also validated by preliminary checks. VOC and CO_2 sensors were grouped together to allow for a more robust comparison between measurements. An infrared people counter was installed at the entrance door to monitor occupant traffic, with a transmitter and receiver pair mounted such that an infra-red beam was interrupted when occupants pass through the door, (and a count was registered). A Q-Scan Twincomm V2.0 Twin Beam Counter manufactured by Axiomatic Technology Limited was used to validate occupancy patterns, logging taking place at its minimum (30 minute interval). While the counter is unable to detect multiple occupants crossing the infrared beam, entrance width should preclude double counting. Figure (3.2) shows the placement of sensors in this test area.

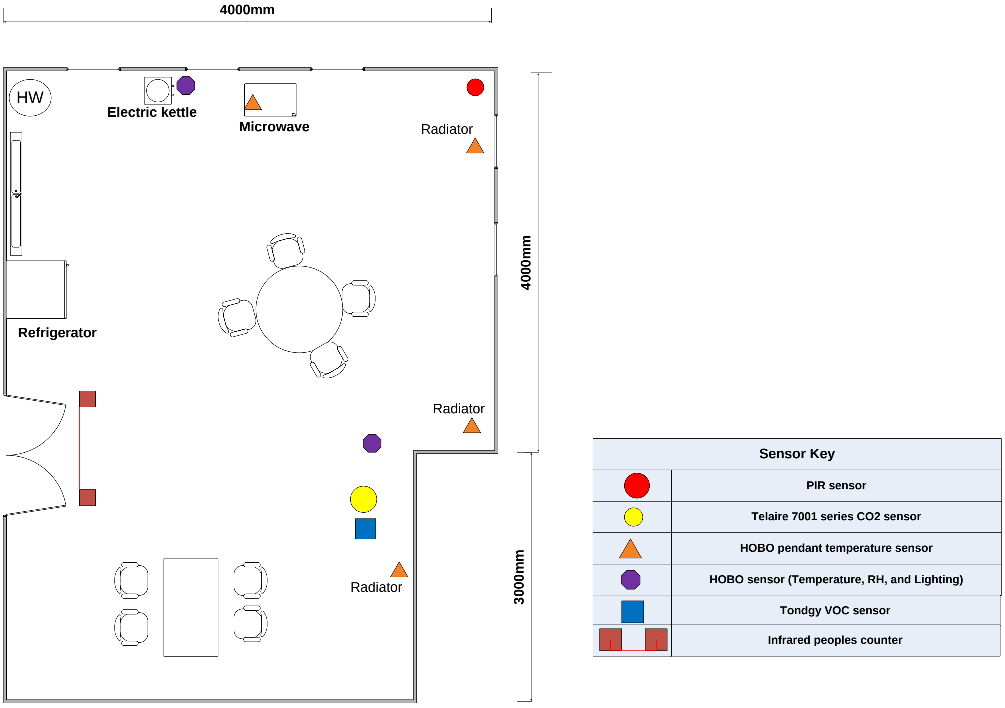


Figure (3.2): Sensor placement in test area one

- **Test area two**

Test area two is the space occupied by a research group providing accommodation for about 18 fulltime PhD research students and 21 staff members. The area comprises of an open floor area, a kitchen, printing bay, equipment room, MSc area and 4 office rooms. About 40 desktop computers and 5 printers are placed at different locations within this space. This space is an open-plan office with high ceiling, and there is a hallway between cubicles accommodating PhD students and some staff members. Figure (3.3) provides a clearer description of the space.

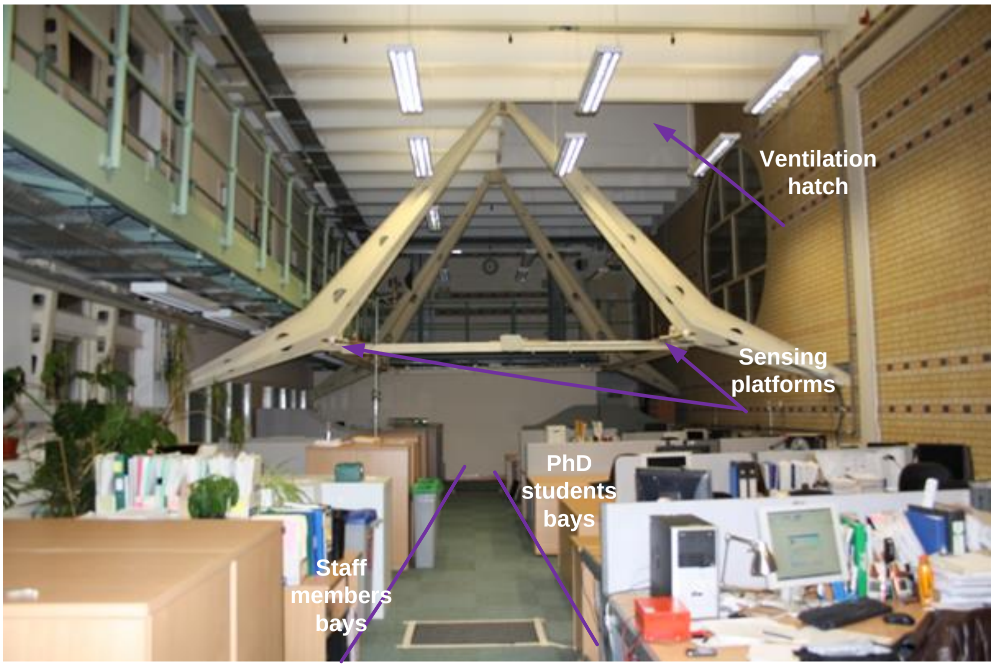


Figure (3.3): IESD Open-plan space

The area enjoys good natural lighting due to its large side windows, although it is shaded from the direct effect of the sun by an adjacent part of the building. It is naturally ventilated with three glazed roof vents. The vents are controlled together by room temperature sensors. The stack or roof vents open up to introduce natural ventilation to cool the space. If the internal temperature is 4°C greater than the set point, then the stack dampers are opened. If the wind speed exceeds a 20 mph then the stack dampers/motorised windows are closed.

Heating in the Queens building is provided by three gas fired boilers, one condensing boiler and two high frequency boilers. However, this test area is heated by radiators fed from one pipe loop from the group of three boilers used. The radiators and three roof vents in the open plan section of the space are placed on just one side of the room (along the cubicles accommodating PhD students). Stack effect in this space is useful in ensuring adequate natural ventilation. However, it has the potential of causing air infiltration in to the space and therefore presents its own challenges to the monitoring strategy implemented in the research.

In an attempt to ensure thoroughness of the monitoring campaign, the main floor was divided into several zones, which do not have any physical boundaries, and therefore, environmental conditions in the zones could interfere with one another, see figure (3.4) for illustration. The MSc area has been excluded from the campaign as it is rarely occupied. The nature of the test room in terms of occupancy and its design layout makes it a representative monitoring region.

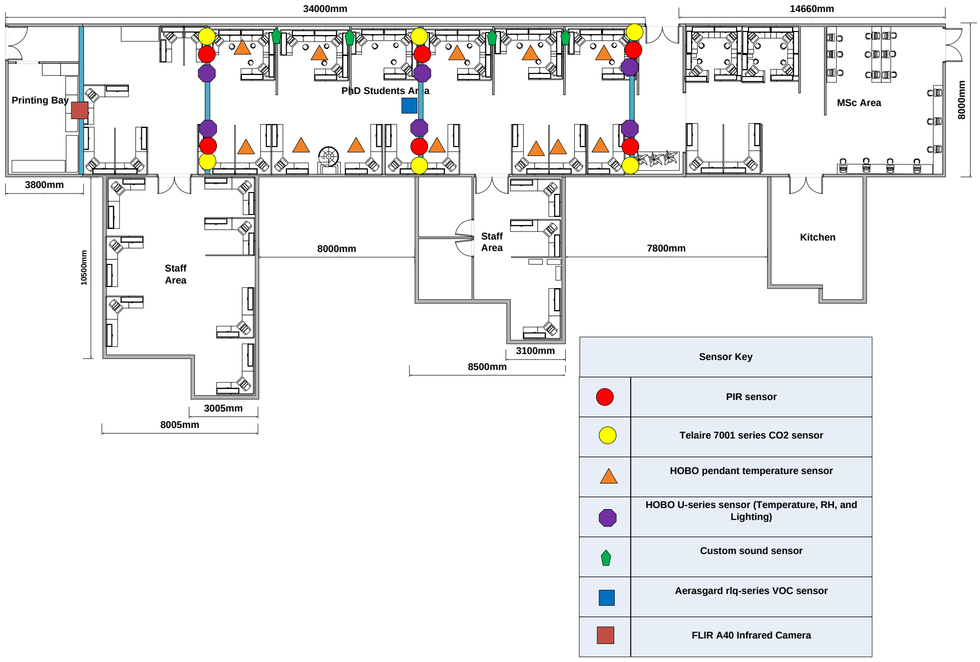


Figure (3.4) Sensor placement in test area two

In test area two, 6 sensing platforms were deployed with each consisting of sensors monitoring indoor environmental parameters such as temperature, lighting, relative humidity, CO_2, and motion, while VOC levels were monitored by a single sensor. Figure (3.5) shows a typical sensing platform in test area two. One sensing platform has been placed at the four corners of each zone in order to increase the sensitivity of PIR sensors, based on a similar approach had been utilized for monitoring occupancy pattern and health status of elderly people (Kusiak et al., 2010). Temperature sensors were attached to the case of desktop computers, while sound sensors were placed close to occupants as depicted in figure (3.4). A thermal infrared camera was placed at a central position to obtain real-time occupancy information which was used to validate results.

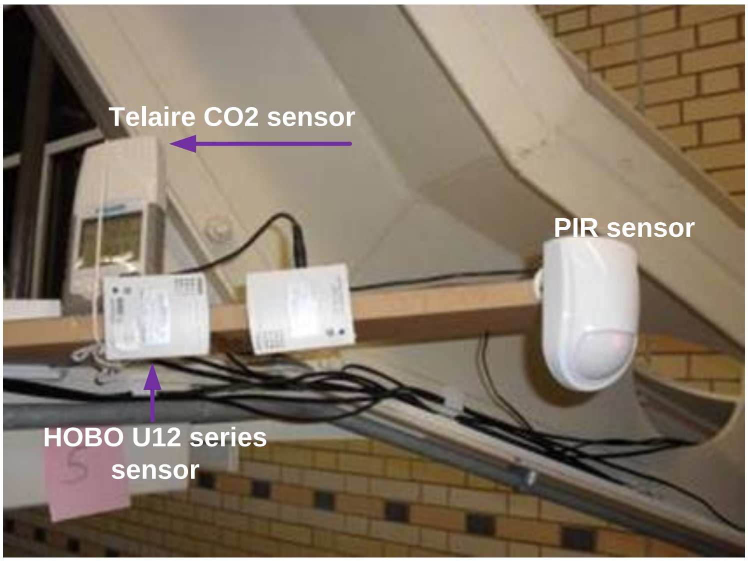


Figure (3.5): Typical sensing platform in test area two

- **Test area three**

The third test area is an admissions office within the Queens building. It is also an open-plan office which accommodates 6 members of staff. It has a small kitchen with fairly high humidity levels, capable of containing about two persons at a time. There are 6 desktop computers and a printer in the space. The room has one exit door, high ceiling and a large rear window that is usually kept locked, although there are smaller windows at the side which are often put to use by occupants. Being an admissions office, both staff and students frequently enter the space to make enquiries. Sound activities are also predominant in the space, with marketing materials being prepared. Airflow rate in this room is less compared to test area two, and as a result there is build-up of indoor parameters. Same indoor thermal conditions as in test area two are maintained for this space. Figure (3.6) shows the instrumentation for this test area.

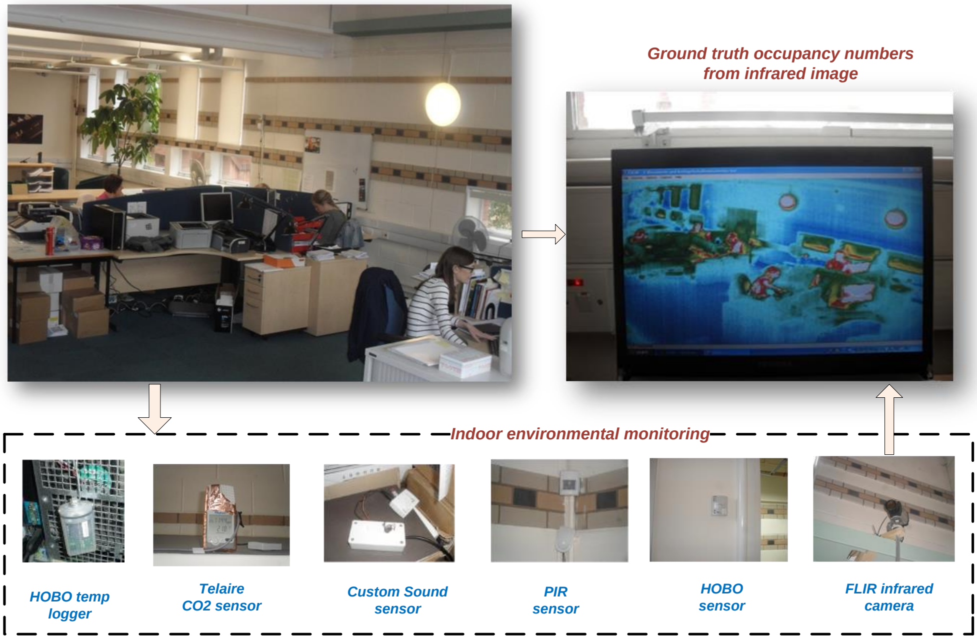


Figure (3.6): Test area three instrumentation

In test area three, PIR sensors were placed close to the existing ones (used by BEMS) to monitor motion, HOBO temperature sensors were attached to the case of all six desktop computers to infer usage pattern, while self-contained HOBO U series dataloggers were employed to monitor the indoor climate, including temperature, humidity, and illumination. VOC and CO_2 levels were monitored using Aerasgard rlq-series air quality sensor, and four GE Sensing Telaire CO_2 sensors respectively, with results being logged using HOBO dataloggers. Ambient sound levels were monitored using custom designed circuitry, which would record as an event sound level over a preset threshold. All events were recorded using HOBO event loggers. Figure (3.7) shows the placement of sensors in test area three.

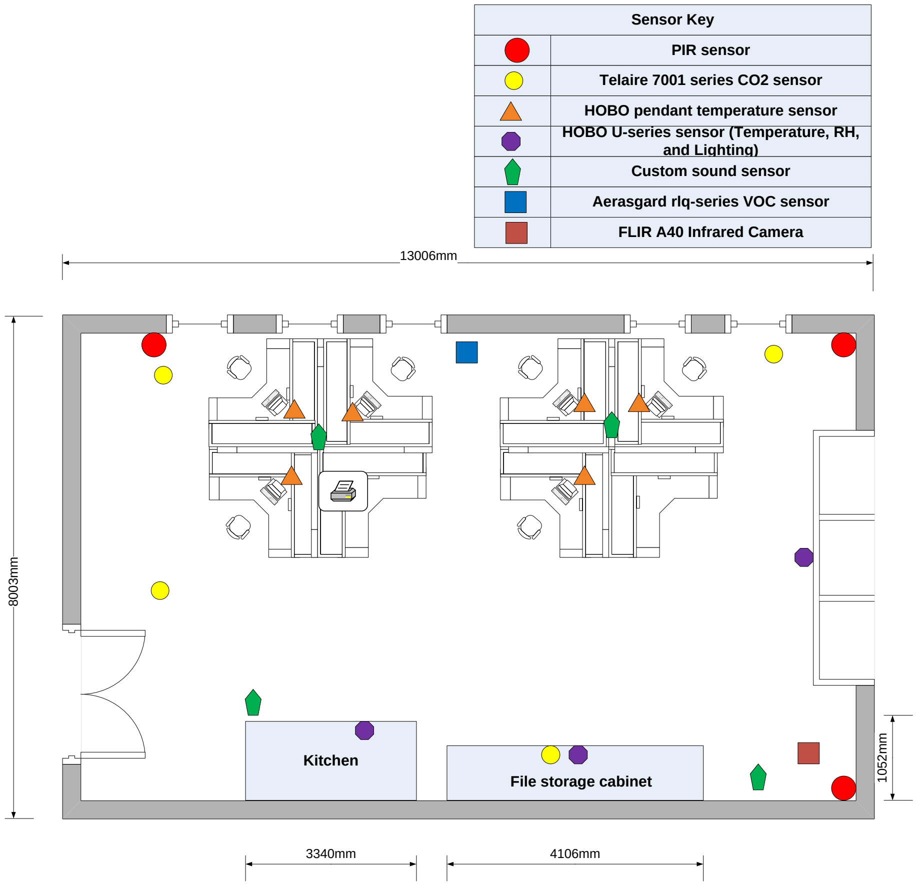


Figure (3.7): Sensor placement in test area three

CO_2 sensor has been placed within the personal work space of occupants in the literature (Hailemariam et al., 2011), which may not mostly reflect contemporary industry practice (CIBSE, 2009), since it may be too intrusive. However, in this research a different placement rationale have been utilized. CO_2 and VOC sensors were placed at reasonable distance from CO_2 and VOC sources respectively. Sound sensors were placed in various positions such that they could capture multiple sound events.

3.2.2 Sensors and observed variables

The sensors of interest in this study were the ones that measures parameters considered to be linked with occupancy and at the same time provide information about the indoor environment under test, which may also be useful for other applications (e.g. IAQ). Figure (3.8) shows the scope of instrumentation used in the research. Individual sensors monitor sound levels, carbon-dioxide, air temperature, relative humidity, desk-top computer usage, VOC, motion, illumination, footfalls, energy (electricity), and outdoor climatic variables.

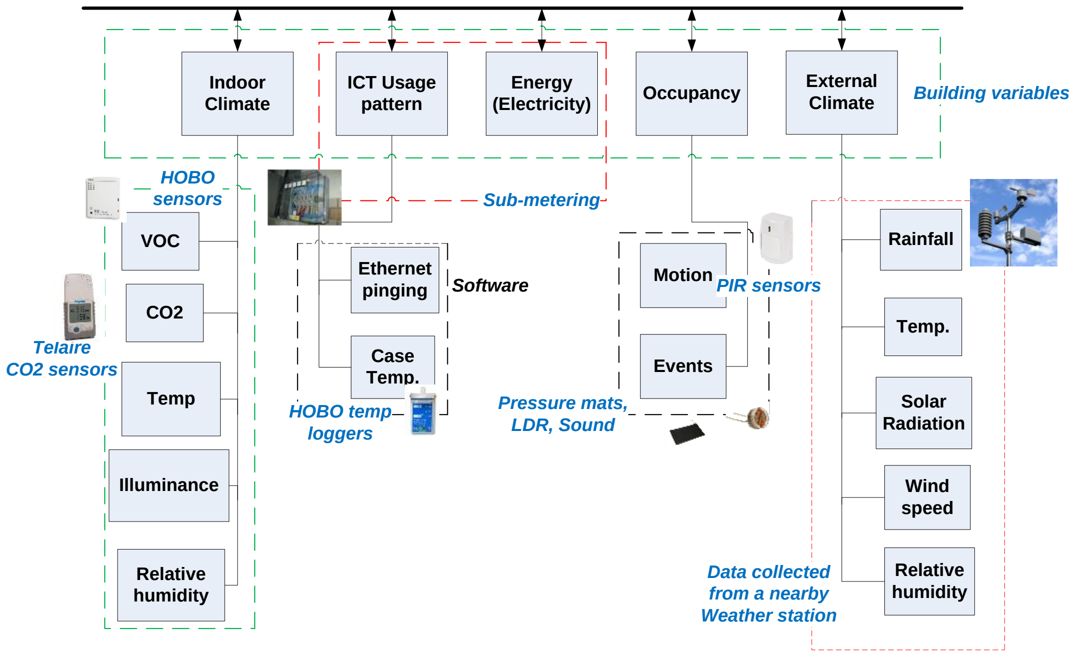


Figure (3.8): Instrumentation scope for the research

Sensors selected in this study offer solutions to the following issues: Intrusion, cost, ease of installation, power consumption, and computational requirements. Ideally, the sensors should have the following properties;

- The sensors should be perceived as non-invasive.
- Sensor data collected should be anonymous, and thus not reveal sensitive information.
- Sensors should be low-cost and preferably available off-the-shelf technologies.
- Sensors should be easy to install, such that it is portable and does not require any specialist training for deployment.
- Sensors should produce data that require minimal computational resources (e.g. a desktop computer) for processing.

- Sensors should require low-maintenance, such that they are easy to replace and maintain (i.e. calibration and quality assurance considerations), and robust to damage.
- Power consumption of sensors should be considered reasonable, low and capable of running for as long as possible on readily available batteries. Figure (3.9) shows stages in the monitoring process.

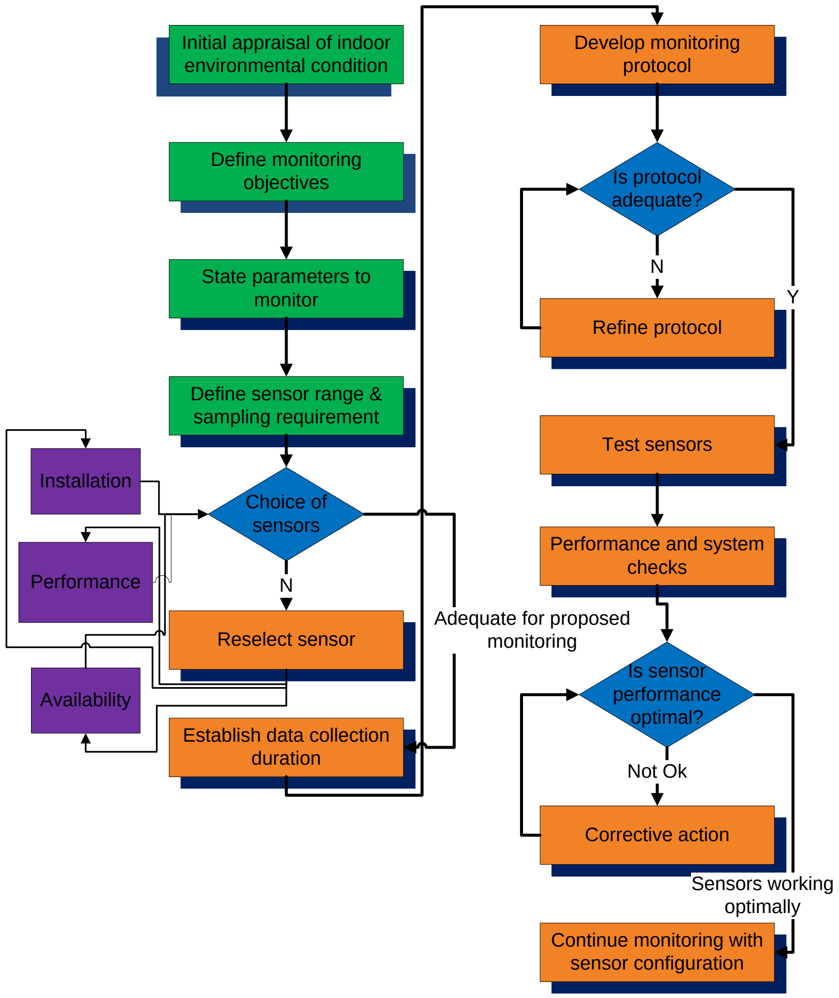


Figure (3.9): Stages in the monitoring process

Hand held instrumentation was initially used to obtain random samples of selected variables at various times of the day, the expected peaks and minimum, to get a tentative range of values for equipment selection and sampling requirements. The monitoring time frame adopted for data collection purposes was largely informed by studies in the literature review. Collecting data over several weeks is representative

(Yang et al., 2012), (Dong et al., 2010), and this will be shown to be effective for occupancy levels estimation in chapter five. Sensors were selected based on requirements presented in Table 2.2. However, availability was the chief motivation for selection. Procedures for operating sensors as per manuals were followed, before sensors were test ran. Initial sensor measurements were checked for reliability, and corrective actions taken where necessary.

- **Monitoring protocol**

The monitoring protocol specifies all the sensors used in the experiment, and the way they were deployed, including data handling aspects. The following protocols have been added to the appendix section;

- Procedures for launching data loggers
- Procedures for data retrieval from sensors
- Procedures for data transfer to a MySQL database

- **Indoor climatic monitoring**

Temperature, relative humidity (RH) and illumination logging utilised a 12-bit combined HOBO sensor (U12-012). Sensor resolution is 0.03^0C at 25^0C, range of -20^0 and 70^0C, accuracy of ± 0.35°C from 0° to 50°C for temperature measurements, while for RH measurements, sensor resolution is 0.03%, accuracy of ± 2.5% for readings between 10 and 90% RH, and range of 5% and 95% RH. Finally, it has a measurement range of 0 and 32300 Lux. The sensor has a sampling rate of between 1 second and 18 hours, user selectable (Onset-Corperation).

Telaire 7000 series CO_2 sensors manufactured by GE sensing were utilised. These use NDIR technology for CO_2 detection, with a display resolution being ±1ppm and repeatability ±20ppm. Response is 60 seconds for 90% of step change, with accuracy of ±50ppm or 5% of the reading whichever is greater. The temperature dependence is ±0.1% of reading per ^{0}C or ±2ppm, whichever is greater, referenced at 25^0C. CO_2 measurements were logged using a HOBO U12 series logger (GE-Sensing).

The range of PIR sensors used is up to 12m, and has an angular coverage of 110^0, operating temperature of between 0 and 50°C, sensor output voltage data being logged using a HOBO U9-001 state logger.

The F2000TSM-VOC sensor manufactured by Tongdy Controls Limited was used to monitor VOC levels in test area one. This sensor has an analogue linear output of 0-10V, although custom electronic circuitry was used to adjust the output voltage to 0-2.5V, being the input voltage range of an HOBO U12 series datalogger. The Aerasgard RQL-series air quality sensor was used for VOC measurements in test area two and three, due to power circuit failure of the F2000TSM-VOC sensor. The Aerasgard RQL sensor has an output of 0-10V/4-20mA, accuracy of ± 20% of final value, operating temperature of between 0 and 50°C. Both VOC sensors deployed for data collection in this work uses a tin metal oxide sensor for the detection of VOCs.

Table (3.1) Indoor climatic monitoring

Parameters monitored	**Equipment used**	**Model**	**Manufacturer**
Carbon(iv)oxide	CO_2 Sensors	7001 series	GE Sensing
Temperature, Relative humidity, illumination	HOBO dataloggers	U12-012	Onset Corporation.
VOC	Air quality sensor	RLQ series & F2000TSM	Aerasgard & Tongdy Controls Limited respectively
Sound level	Sound sensor		Custom design by researcher
Motion	PIR sensor		Maplin-Retailer

- **Appliance usage monitoring**

Electronic equipment usage was monitored using case temperature measurements. Most electronic equipment such as a desktop computer dissipates heat when switched on and its case gets warm, when power is flowing. Similarly, the case temperature drops to ambient conditions when switched off, or not using electricity.

The case temperature normally reaches few degrees (between 2-8^{0}C) higher than the surrounding air temperature. Previous research (Brown and Wright, 2008, Brown et al., 2011), have shown the use of this technique is an easy way to gather information on appliance usage, although it has not been utilised for occupancy numbers estimation in office buildings. After initial pre-processing to remove noise from the case temperature measurements, a rule based processing was used to establish appliance duty cycles; this technique is presented in more detail in (Brown and Wright, 2008). To monitor duty cycles of computers, a pendant HOBO UA-001-08 logger was attached to the case of an electrical appliance, adjacent to the warm air from ventilation slots, as in figure (3.10). The logger has the same specification as the HOBO U12-012 for temperature measurement.

Figure (3.10): Case temperature monitoring

Table (3.2) Appliance usage monitoring

ICT Equipment	**Parameter**	**Monitoring Device**	**Manufacturer**
	Case Temperature	Pendant HOBO temperature sensor	Onset Corporation

- **Outdoor climatic monitoring**

Weather data, including variables such as rainfall, air temperature, solar radiation, wind-speed, direction, relative humidity and barometric pressure were obtained from a weather station within the university campus.

Table (3.3) Outdoor climatic monitoring

Parameters Monitored	Equipment used	Manufacturer
Rainfall	Rain-gauge	Delta-T devices limited
Air Temperature	Thermistors	Delta-T devices limited
Wind speed & Direction	Wind vane and Anenometer	Delta-T devices limited
Solar irradiation	Sunshine sensor & Pyranometer	Delta-T devices limited
Barometric pressure	Barometric Pressure sensor	Delta-T devices limited

- **Energy Monitoring**

Energy (electricity) monitoring was carried out using existing sub-meters in the building, provided by Energy Metering Technology Limited, see figure (3.11).

Table (3.4): Energy monitoring

Parameter	Equipment used	Manufacturers
Energy (Electricity)	Sub-meters	Energy Metering Technology (EMT) Limited.

Figure (3.11): Existing sub-meters in the test building

3.2.3 Data collection and Connectivity

Data collection (weekdays and weekends) was in five continuous periods from the three test areas. Table (3.5) provides details of each dataset. Weather data were sampled every 10 minutes, with hourly averages logged. Half-hourly electricity data were relayed via a low-power radio network to a central receiver, and then uploaded to a *MySQL* database server. Indoor environmental sensor data were downloaded to a PC using a HOBO shuttle and uploaded to a *MySQL* database using *MATLAB* scripts. Figure (3.12) illustrates the relevant data connectivity flows. Occupancy estimates from the data were subsequently analysed in *MATLAB* and *WEKA*.

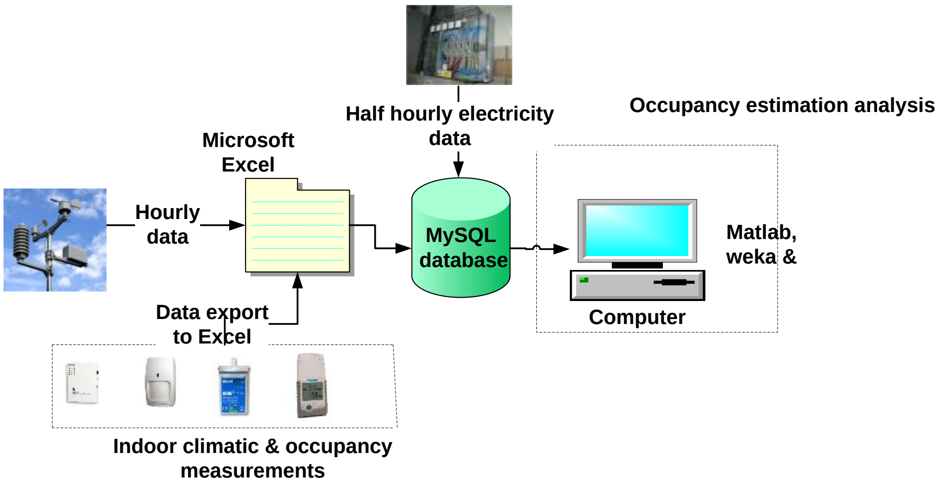


Figure (3.12): Data connectivity flow diagram

Table (3.5): Data collection periods for indoor environmental measurements

Test area	**Duration**	**Data type**	**Sampling interval (minutes)**
		Group One	
One	02/12/2010 00:00:00 – 16/12/2010 00:00:00	CO_2	2
		Kettle RH	10
		Microwave case Temperature	10
		VOC	2
		Radiator temperature	10
		Ambient temperature	10
		Illumination	10
		Lighting case temperature	10
		Ambient RH	10
		Motion	1
		Group Two	
Two	07/07/2012 08:00:00 – 08/08/2012 00:00:00	CO_2	1
		VOC	1
		Sound	1
		Ambient RH	5
		Ambient temperature	5
		Desktop PC case temperature	5
		Illumination	5
		Motion	1

Group Three			
Three	01/07/2012 00:00:00 – 18/07/2012 15:30:00	CO_2	1
		VOC	1
		Sound	1
		Ambient RH	5
		Ambient temperature	5
		Desktop PC case temperature	5
		Illumination	5
		Motion	1
Group Four			
Three	12/09/2012 00:00:00 – 11/10/2012 00:00:00	CO_2	1
		VOC	1
		Sound	1
		Ambient RH	5
		Ambient temperature	5
		Desktop PC case temperature	5
		Illumination	5
		Motion	1
Group Five			
Three	27/11/2012 11:00:00 – 20/12/2012 23:00:00	CO_2	1
		VOC	1
		Sound	1
		Ambient RH	1
		Ambient temperature	1
		Desktop PC case temperature	1
		Illumination	1
		Motion	1

3.3 Pilot experiment

The purpose of the pilot experiment was to provide an insight in to sensor selection useful for occupancy monitoring. Data collection from test area one was used as the pilot experiment. In the next section, initial findings are presented.

3.3.1 CO_2 and VOC measurements

In Figure (3.13), VOC concentration was fairly constant between 00:00am and 05:00am throughout the weekdays, although outdoor temperature and relative humidity may have been an influencing factor. Intermittent peaks in the data were quite visible from about 08:00am with fairly sporadic occupancy following until around 12:00pm -5:00pm, when traffic became heavy. Occupancy count was usually more than one and VOC concentration levels peak, between 1.35 and 1.6, indicative of the concentration levels. Occupants mostly had their lunch during this period, some persons sit and eat in the kitchen while others just microwave their food and vacated the space.

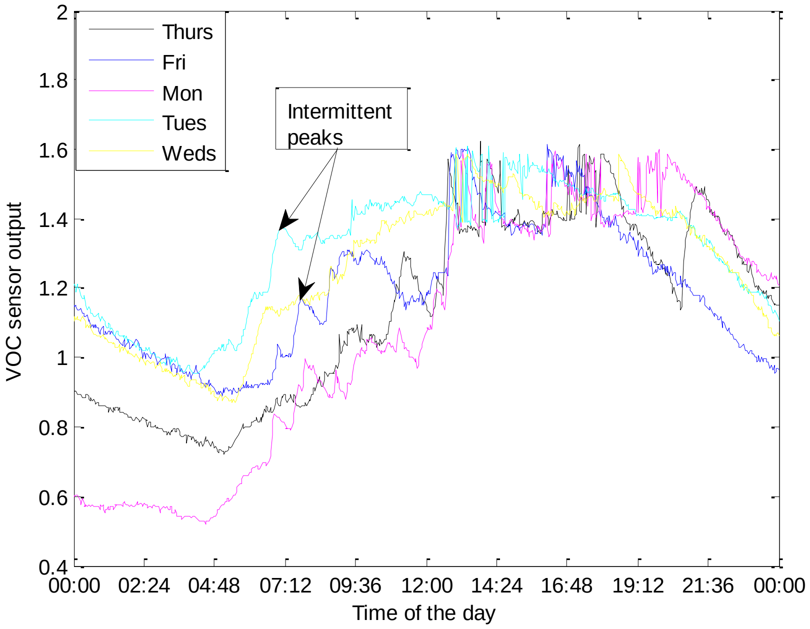


Figure (3.13): Daily VOC measurements

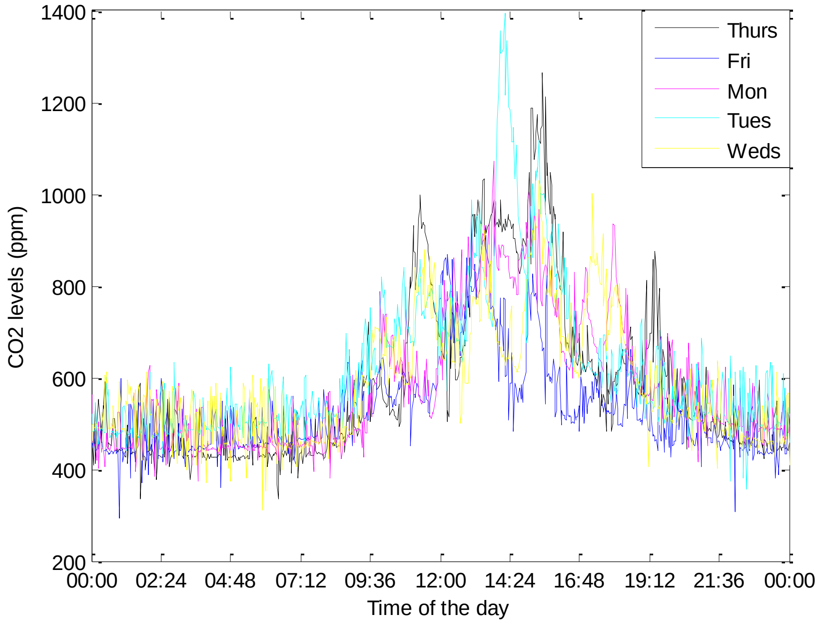


Figure (3.14): Daily CO_2 measurements

The trend in CO_2 data was similar to VOC data for weekdays especially from 12:00pm to 16:48pm, CO_2 varied between 800 -1400ppm as shown in figure (3.14). Concentration after work was fairly constant between 400ppm and 600ppm. Between 8:00am and 12:00pm (during this time space was expected to be occupied at least for short periods). Occupancy events were not obvious for CO_2 measurements compared to VOC, which is known to respond more quickly. However, between 12:00pm and 16:48pm, peaks in the CO_2 levels implying occupancy were clear, since occupancy periods tend to be longer during this time of the day.

Analysing VOC and CO_2 rates of change revealed how both track with occupancy entropy. Using the results from a typical weekday as in figure (3.15), VOC measurements clearly tracked occupancy more effectively than CO_2 measurements in this environment. Unoccupied VOC emissions settled to around 0.02 every 2 minutes. Before 12:00pm, VOC peaks tend to track transient occupants activities. VOC measurements were able to track closely against occupants' activities especially during lunch time (between 12:00pm and 16:48pm). Unoccupied CO_2

levels however were not stable enough to determine a threshold, which suggest significant noise interference may be in the data.

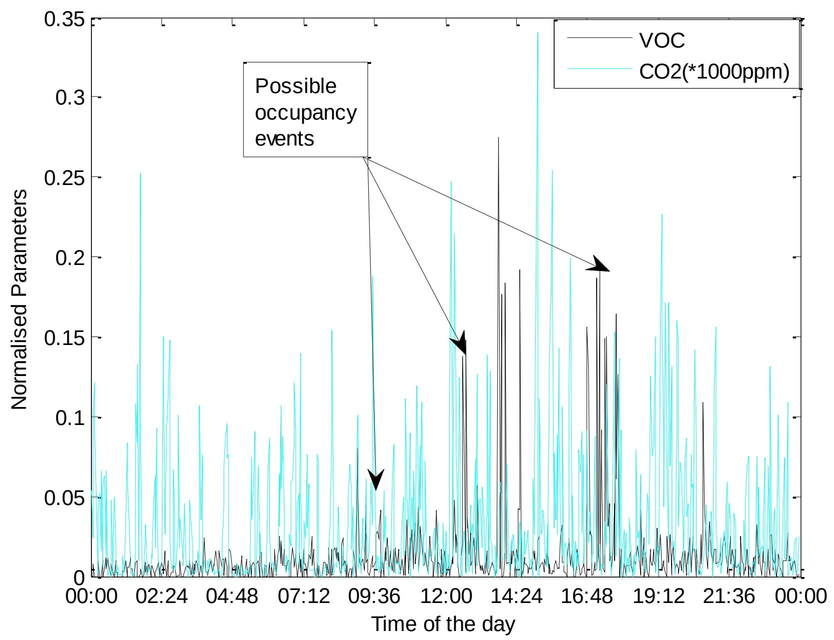


Figure (3.15): Change in CO_2 and VOC measurements

3.3.2 Case temperature measurements

This parameter was notable for providing good indication of occupancy during lunch periods between 12:00 and 16:30pm for weekdays. Figure (3.16) show case temperature measurements over typical week. Case temperature may be up to 8^0C above ambient during lunch time, indicating heavy appliance use. Figure (3.17) suggests that a case temperature change greater than 0.5^0C in 10 minutes indicated device use, and hence occupancy. Clearly, case temperature measurement alone to indicate occupancy requires a high probability that occupants always use appliances (as may be expected in a kitchen). Figure (3.17), between 16:48pm and 19:12pm, illustrates an exception, showing that results may require some caution.

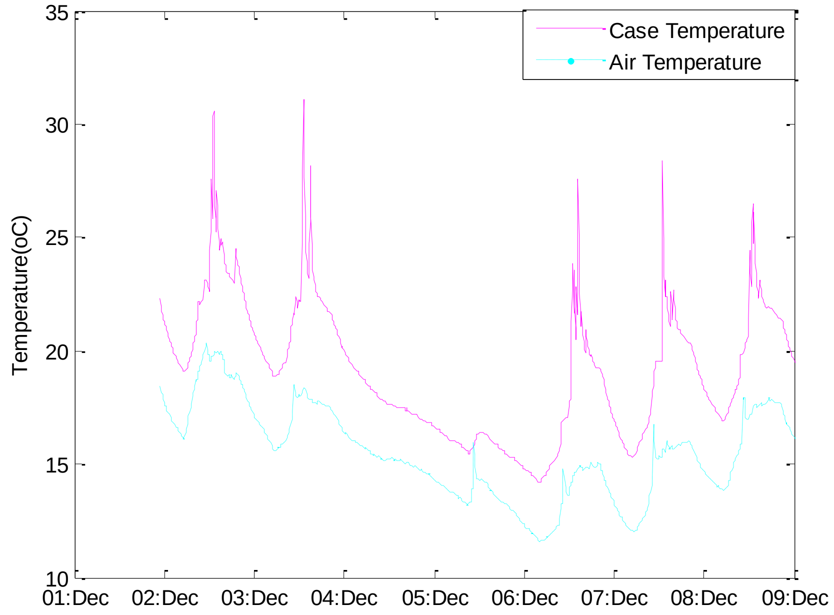


Figure (3.16): Weekdays ambient and case temperature measurements

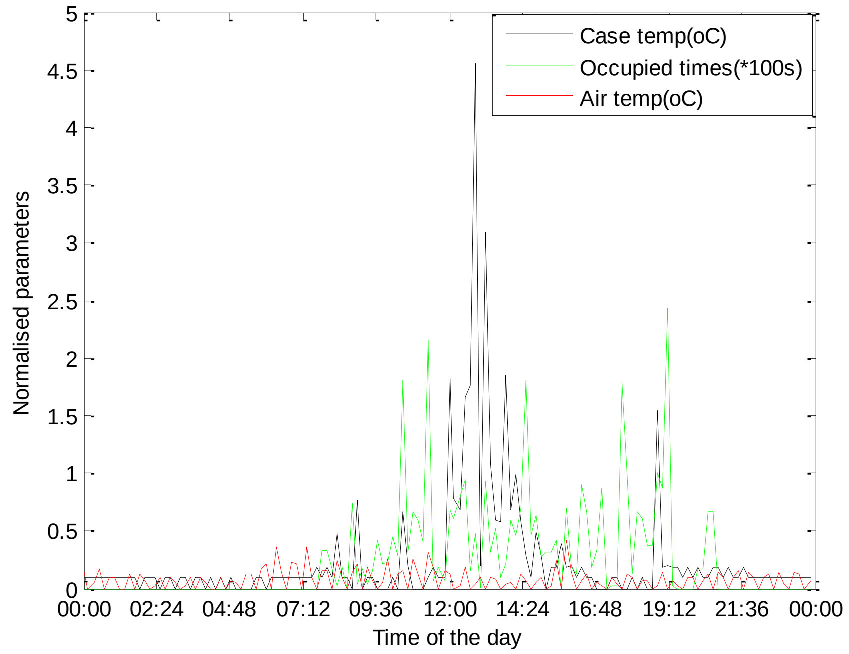


Figure (3.17): Change in case temperature measurements on a typical day

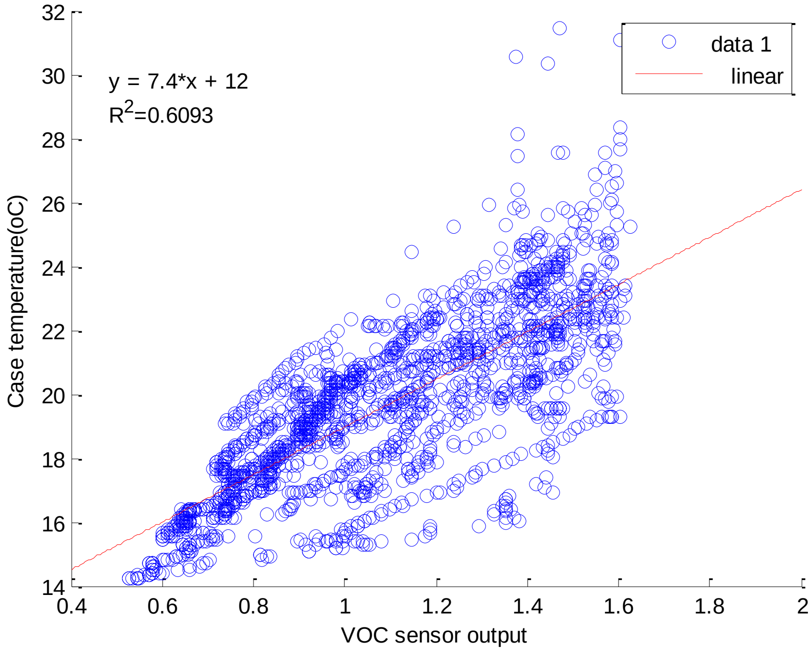


Figure (3.18): Correlation between case temperature and VOC measurements

A scatter plot of the case temperature against VOC showed a good link between both, figure (3.18) also shows reasonable clustering of data with an R-squared value of 0.6093. This adds some weight to the idea that VOCs levels may be a good indicator of occupancy in this setting, although a better R-squared value may establish this notion.

3.3.3 Occupancy validation using infrared peoples counter

After 3 days occupancy was logged at -2, this was not surprising as a people counter device is unable to detect multiple occupants crossing its infrared beam at a time, and routinely fails to return a zero sum at the end of each working day. Hence, results beyond 04/12/2010 were not used for validation. Figure (3.19) shows an increasing measured parameters during occupancy, which would fall naturally outside these times. Between 12:00pm and 17:00pm, measured parameters and occupancy both peak, occupancy peaking at 5. Differences between weekend (04/12) and weekday occupation are clear from figure (3.19). Detection of occupancy entropy was not entirely clear using a 30 minute sample interval (figure (3.20)), and would certainly benefit from higher resolution. However, close inspection of figure

(3.20); still suggests that indoor parameters tracked well with occupancy entropy especially between 12:00pm and 16:48pm.

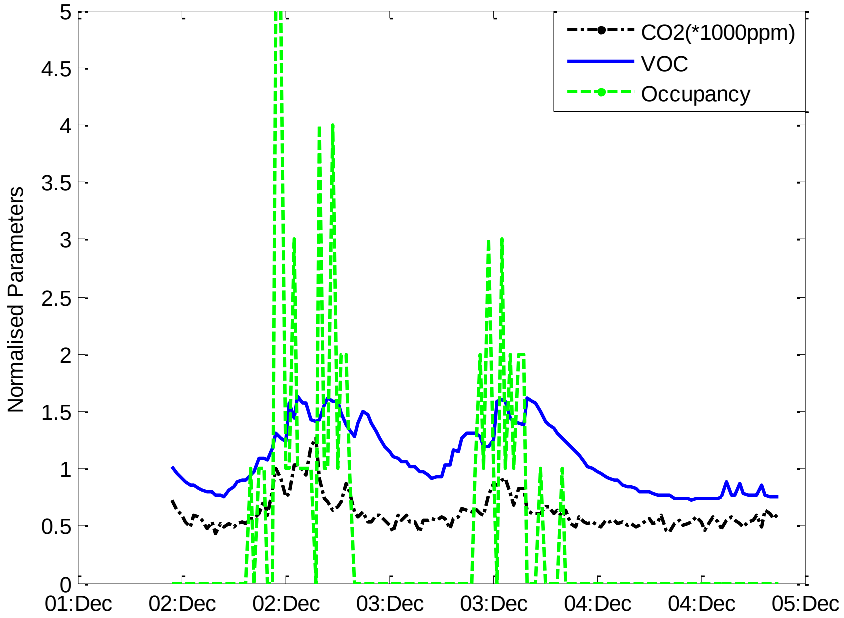


Figure (3.19): Occupancy count and indoor climatic variables

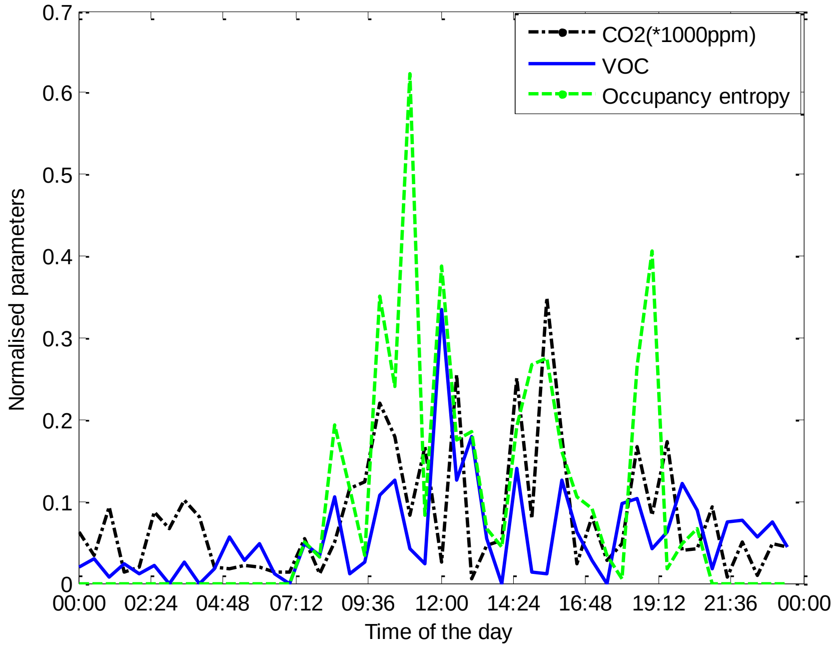


Figure (3.20): Change in occupancy and indoor parameters

3.3.4 Findings from pilot experiment

- The use of a people counter for occupancy validation may not be a reliable method due to over or under counting. A 30-minute sampling period may be too coarse for occupancy numbers tracking.
- Appliance case temperature monitoring may be a useful way from which occupancy can be inferred, especially when if appliances under test are regularly put to use, otherwise, it may be limited for occupancy numbers.
- VOC measurements may be useful for occupancy tracking, in environments where VOC sources are prevalent such as an office kitchen setting.
- CO_2 measurements for occupancy tracking were inconclusive. Temporal analysis to assess CO_2 level changes with occupancy entropy did not produce reasonable results, as levels were never stable. This may suggest the presence of noise in the measurements, or calibration issues.
- The results may be limited to test area one, as activities, and therefore indoor climatic conditions may be different compared to that in test areas two and three.

3.4 Carbon dioxide (CO_2) sensor refinement

In this section, initial challenges with CO_2 sensors during data collection and the modifications carried to improve CO_2 measurement integrity are presented.

3.4.1 Challenges with initial CO_2 measurements

The findings from the pilot experiment suggested noisy CO_2 measurements. Calibration tests were initiated to ascertain the measurement integrity of CO_2 sensors. A zero calibration procedure using nitrogen gas was carried out on CO_2 sensors before being deployed for data collection as illustrated in figure (3.21). Although, sensor display showed 0ppm, no one sensor under test recorded a CO_2 concentration level of 0ppm, as per results retrieved from data loggers. This was not surprising, since CO_2 sensors are known to suffer significant drift over time, as mentioned before (section 2.4.6). These sensors were acquired in 2007, while the calibration test was carried out in 2011. Sensor drift associated with these sensors was in the region of ± (23-195) ppm. Table (3.6) shows the average drift for some sensors after zero calibration.

Figure (3.21): CO_2 sensor in a zero calibration mode

Table (3.6): Average drift of some sensors after zero calibration

Sensor ID	Average drift (V)	Displayed CO_2 concentration (ppm)
IESD_027	0.1860	186.0
IESD_017	0.0546	54.6
IESD_023	0.0237	23.7
IESD_019	0.0237	23.7
IESD_016	0.0554	55.4
IESD_006	0.1956	195.6

From the Telaire 7001 series data sheet, the conversion factor is 1mV = 1ppm

Full scale deflection (f.s.d) = 0 - 4V

Span calibration in this context entails calibrating the CO_2 sensor for a known gas concentration to evaluate its measurement integrity. This was carried out using CO_2 gas of 488ppm, as shown in figure (3.22). However, the calibration process was discontinued since CO_2 levels were never stable enough.

Figure (3.22): CO_2 sensor span calibration

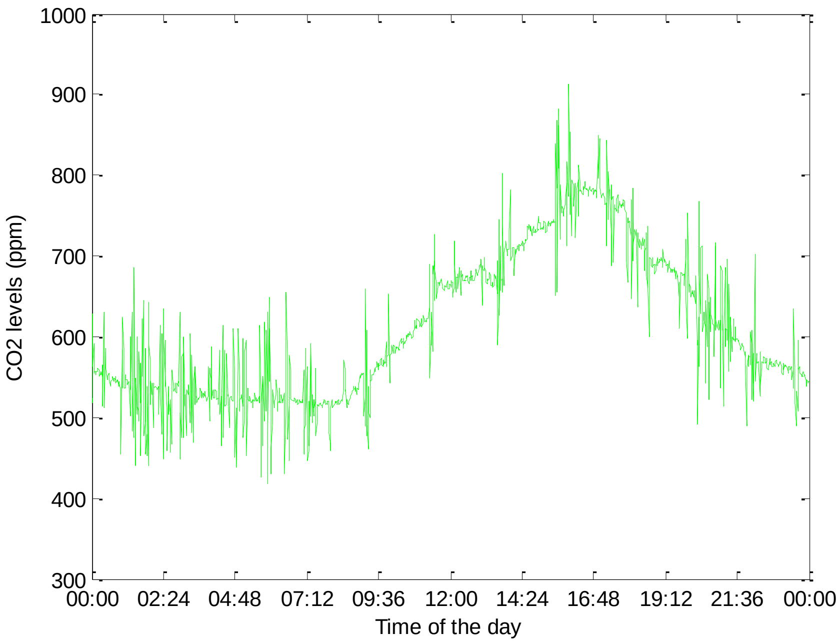


Figure (3.23): Typical daily CO_2 sensor measurement showing EMI

Initial analysis of CO_2 measurements collected suggested the presence of noise, as shown in the figure (3.23). The unoccupied period between 00:00am and 05:00am, showed peaks during period of low (or possibly zero) occupancy, where the concentration levels were expected to be quite stable. The presence of noise in the results informed the need for further investigation to ascertain its source(s).

3.4.2 Electromagnetic interference (EMI) issues with CO_2 power supply

Switch mode power supplies (SMPS) are widely used in the electronics industry, in computers, television receivers, battery chargers etc. SMPSs' generate radio interference due to high frequency switching in internal circuitry, usually above 20 kHz. This interference is propagated through space by means of electromagnetic fields or via the mains supply. This interference usually consists of switching frequency and many associated harmonics, and can lead to performance degradation of electrical or electronic equipment with an installed SMPS or nearby equipment. CO_2 sensors used in this research were powered using a 6-9V mains SMPS.

EMI sources include, but are not limited to mains power hum, FM radio transmitter etc. EMI can be coupled into a circuit through several paths such as transmission wires, cables or printed circuit board (PCB) traces. For EMI to be present, three elements that must exist are a source, coupling means and a receptor (victim circuit). Figure (3.24) shows this illustration, exclusion of any one of these elements will eliminate EMI, in practise this can only be reduced significantly. Effects of EMI in equipment can be minimized by reducing the susceptibility of the receptor, suppressing emissions or reducing the efficiency of the coupling path.

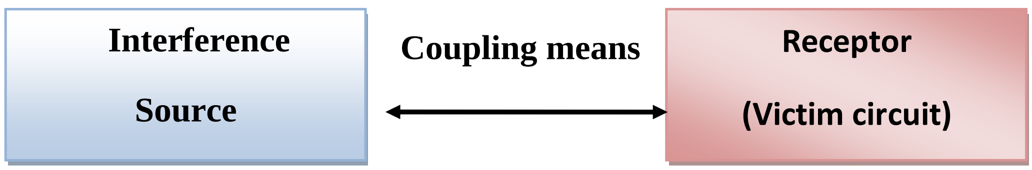


Figure (3.24): Necessary elements for electromagnetic interference

3.4.3 CO_2 sensor modification process

EMI generation can be reduced by proper design of circuit layout and selection of circuit components (Mainali and Oruganti, 2010). These methods are mostly implemented during an SMPS design phase, some of which include proper routing of tracks, proper use of ground planes, power supply impedance matching, and reducing logic frequency to a minimum (Mainali and Oruganti, 2010).
Close inspection of the results as shown in figure (3.23), suggests the presence of continuous harmonics in the sensor measurements. The CO_2 power supply being an

SMPS was suspected to be a potential cause. Other possible radio frequency (RF) sources in the observed environment, which may couple EMI into the sensor's measurement circuit, included an FM student radio station and hundreds of computers, and *WIFI* station or access points. The TR10R060- 12A03 manufactured by (CINCON) was the SMPS used for powering CO_2 sensors. According to the data sheet, the device has good electromagnetic compatibility as per standards (CINCON). Although, sensor measurement results suggested otherwise, with a caveat that the observed environment contains RF sources.

EMI can be transmitted through radiation as electromagnetic waves in free space or by conduction via a conductive means, or both. Various suppression solutions are available, including the use of EMI shield laminates, ferrites (in the form of beads or cable shields), low-pass filters, conductive coatings etc. However, solution such as the use of a copper foil containing conductive adhesive is recommended for EMI shielding of electronic equipment enclosed in plastic enclosures (Devender and Ramasamy, 1997). The use of laminates is less time consuming, and typically provides a reliable ground surface, shielding of 20dB to 60dB depending on frequency, configuration and installation (Devender and Ramasamy, 1997). For this research, emphasis was not placed on re-designing the SMPS.

3.4.4 Experimental set-up for CO_2 EMI mitigation

In order to investigate if the sensor's SMPS was responsible for noisy measurements. Six sensors were selected randomly and placed next to each other in an enclosure as shown in figure (3.25), to allow for robust comparison. Three of the sensors were battery powered while the remaining three were mains powered.

Results in figure (3.26) show that mains powered sensors measurements were more unstable than that of battery powered sensors. The mains powered ones had relatively more spikes than that of the battery powered sensors, possibly due to the switching action from the SMPS, suggesting that it contributed significantly to noise generation in the measurements. Although, battery powered sensors were not completely devoid of spikes.

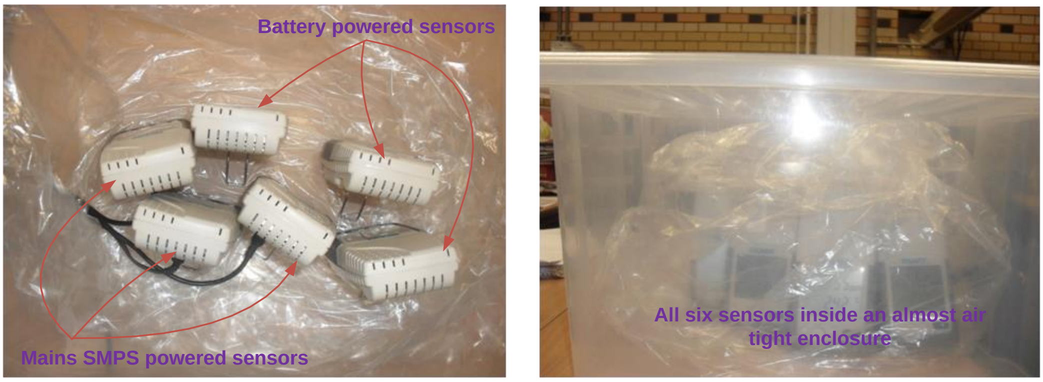


Figure (3.25): CO_2 sensor experiment for EMI investigation.

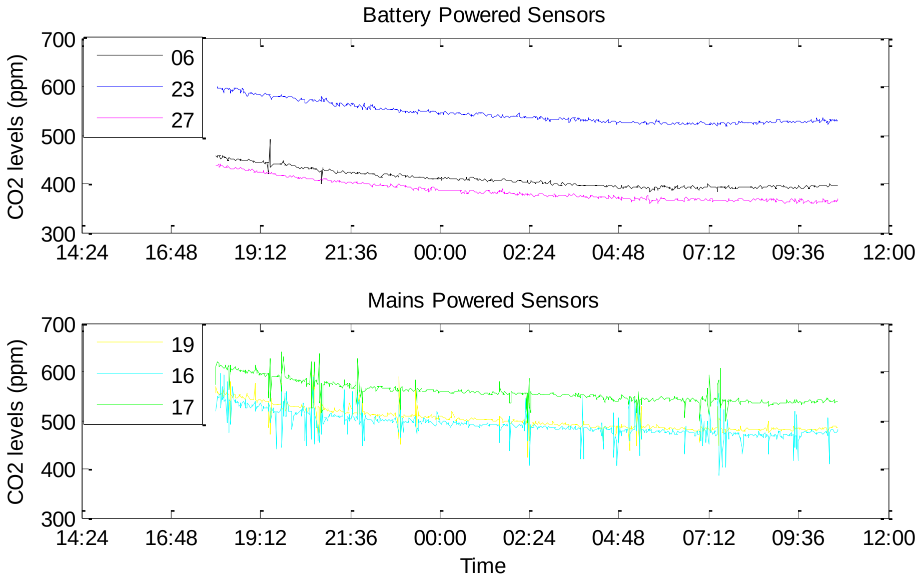


Figure (3.26): SMPS and battery powered sensors test

A Sampling frequency of 16.67Hz, which is 1000 times the rate of data acquisition, was used in a Fast Fourier Transform (FFT) analysis. This was considered sufficient, as per Nyquist criterion. The analysis showed signal strength for SMPS powered sensors to be around 73.80dB/Hz, while the battery powered ones was around 77.73dB/Hz. In an attempt to improve the quality of CO_2 sensor measurements, all sensors were battery powered and shielded using copper foil, before being placed inside a metal office cabinet which was then grounded. Results were visibly clear of spikes, and signal strength increased to around 80.20dB/Hz, as shown in figure (3.27) and (3.28) respectively.

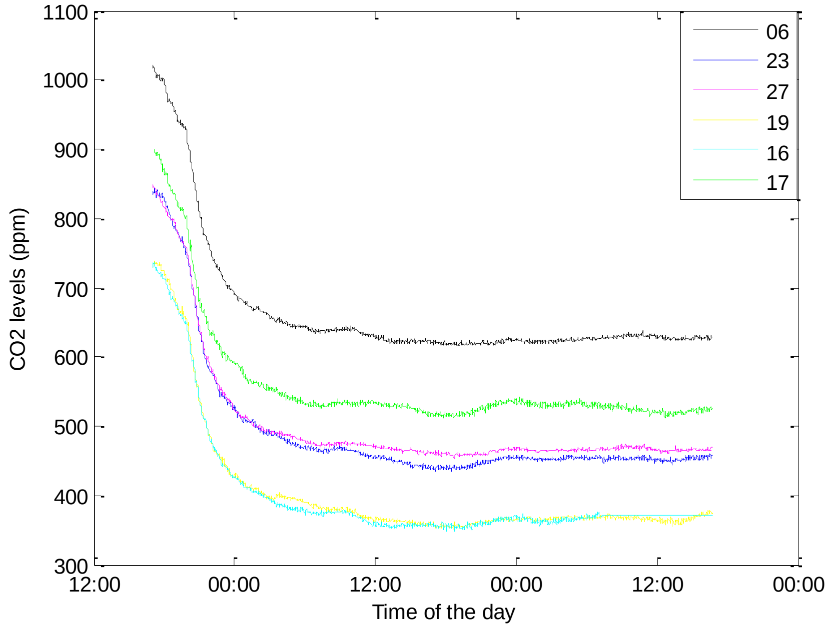


Figure (3.27): All six battery powered sensors showing no spikes

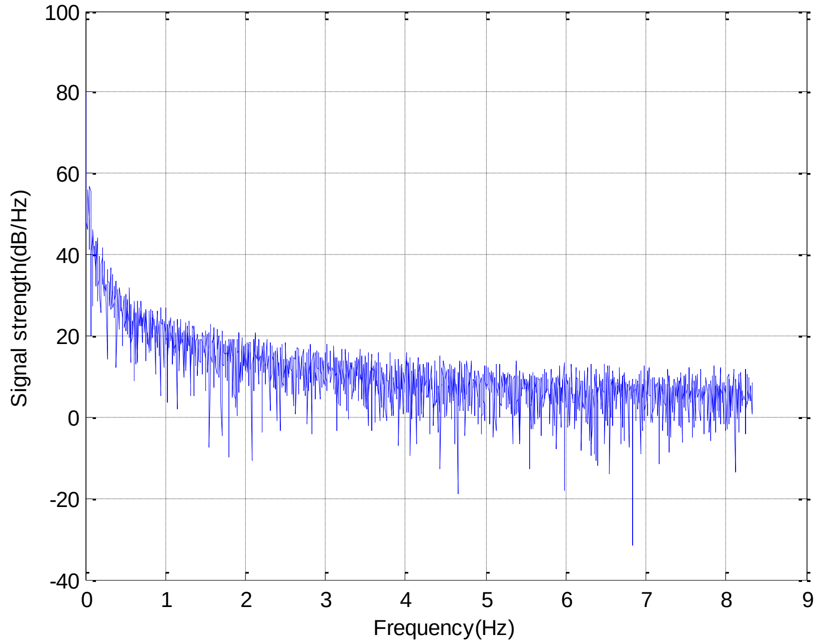


Figure (3.28): Signal strength after shielding

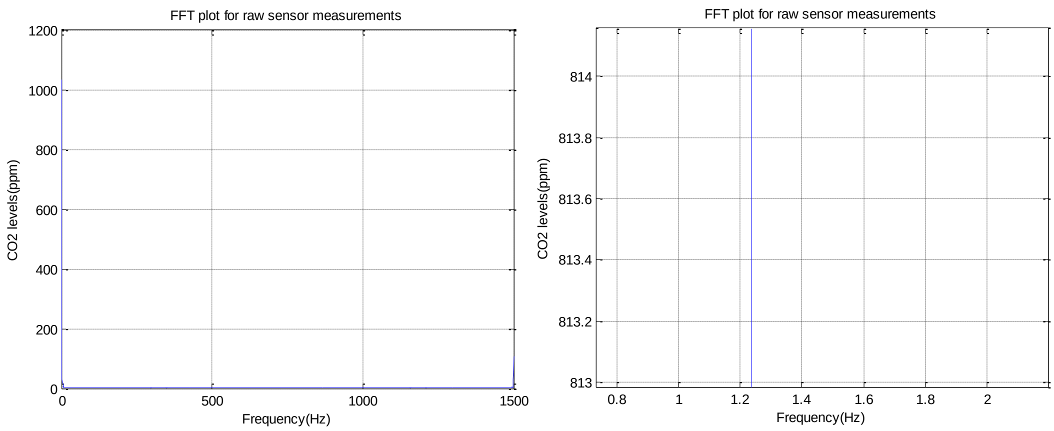


Figure (3.29): FFT plots for unfiltered (raw) signal

Further analysis included applying a low-pass filtering to sensor measurements. FFT analysis showed the signal strength had its highest amplitude at 1.25Hz (figure 3.29); hence a cut-off frequency of 1.24Hz was chosen for filter design. A second-order Butterworth filter was implemented in *MATLAB* and applied to sensor measurements.

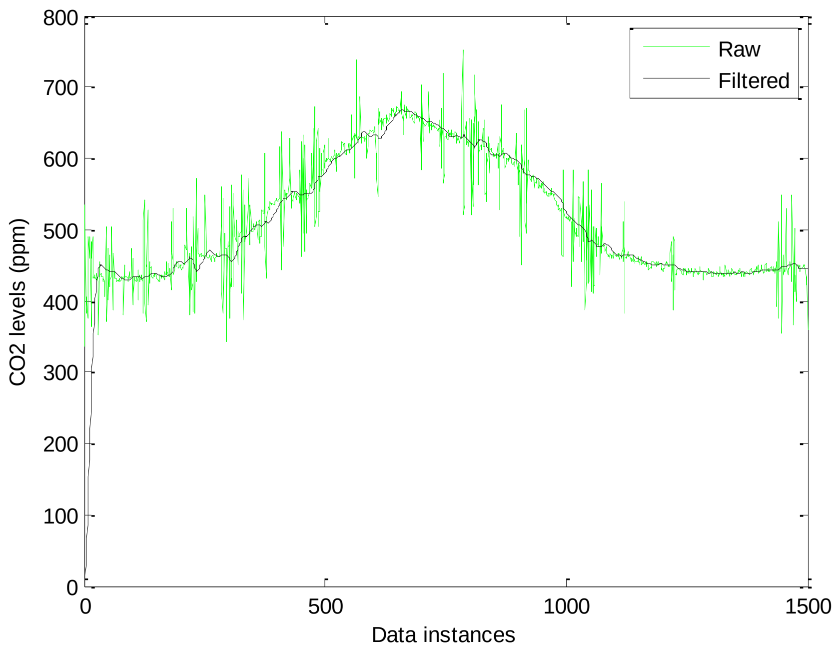


Figure (3.30): Raw and filtered signal

Figure (3.30) shows the results after filtering. An FFT analysis revealed that the filtered signal strength was around 73.75dB/Hz, suggesting the amplitude of the harmonics was reduced, but this may not give a complete representation of the signal, as there may be information loss. For occupancy detection systems based on CO_2 measurements, this may have ramifications. For this research, shielding was implemented with no loops using a copper laminate containing adhesive as part of the EMI mitigation solution, and it is considered simple and effective to implement, with better signal information retention compared to the methods examined. Sensors deployed for data collection were shielded as shown in figure (3.31), and grounded to a 6V 14Ah rechargeable lead –acid battery (which was used as the sensor's power supply).

Figure (3.31): A shielded CO_2 sensor

3.5 Custom sound sensor design

A new low cost sensor was developed and deployed for sound level measurements in this research, see figure (3.32). Outputs from existing sound sensors may normally require some form of data processing such as a FFT analysis, to obtain meaningful information (regarding occupancy). A different methodology for sound level monitoring was employed, one in which outputs from the sound sensor are similar to that of a PIR sensor; it produces a binary output for occupancy and vacancy periods. To the best of the authors' knowledge, no existing sound sensor has been applied in such a way for building occupancy sensing. This section presents the sensor design process as illustrated in figure (3.33).

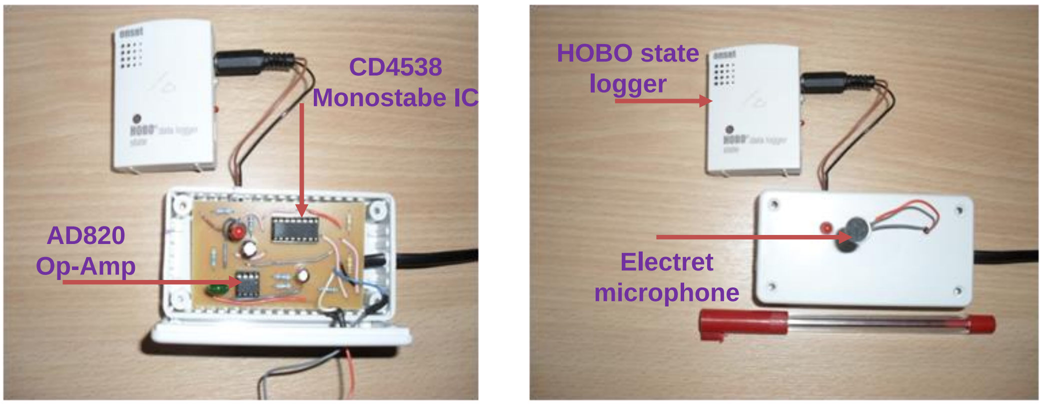


Figure (3.32): Custom sound sensor

3.5.1 Design process

Operational amplifiers are commonly called Op-amps, and are used in signal conditioning processes such as filtering, amplification, etc. They can produce output voltages of several thousand times larger than input signals. Most prominent of their properties include very high open loop voltage gain of about 10^5, very high input impedance typically 10^6 to $10^{12}\,\Omega$, and very low output impedance, commonly 100Ω.

The sound sensor is a simple audio amplifier based on a standard inverting circuit, with sound measurements logged using an HOBO state logger. A key design consideration was to ensure that the sensor output did not exceed the allowable input range of HOBO state logger (0-2.5V), as this could result in electrical loading of the state logger circuitry, which is undesirable. It was intended to sample and store sound measurement once every minute, this information was useful in the time constant design. Table (3.7) presents the design parameters selected.

Table (3.7) Design parameters selected

Gain	100
Feed-back resistor	1MΩ
Input Resistor	10k
Time Constant	100s
Electret microphone sensitivity	1V/µbar at 1 kHz
Input capacitor	10µF
Noise decoupling capacitor	22pF
Maximum operating frequency (*f*)	1kHz
Supply voltage	6V
Diode	2V7
HOBO state logger input voltage	0-2.5V
Microphone input	90mV/Pa

In the circuit of figure (3.34) and (3.35), a single power supply is used and the input voltage to amplified is applied via resistor R_i to the inverting (-) terminal of the Op-amp. The output is therefore in anti-phase with the input. Voltage at the inverting input (point *D*) can never be far from zero because of the high value of the circuits' open-loop voltage gain. Therefore, since *D* is in effect at 0V, the voltage across R_i equals the input voltage, V_i, and that across R_f equals output voltage, V_0. *D* is called a virtual earth (or ground) point, though of course it is not connected to ground. When V_i is positive, current *I* flows as shown through R_i and then through R_f. Only a negligible fraction of *I* enters the inverting input of the Op-amp (partly because of its very high input impedance)

$$G = \frac{V_O}{V_i} = -\frac{R_f}{R_i} \tag{3.1}$$

Where *G* is the Gain of the Op-amp

The time constant stipulates a preset threshold to record sound levels as an event. Usually when there is a sound event, a capacitor, C_t, is charged, and discharged through the resistor, R_t. The duration of this process is dependent on the combination of C_t, and R_t.

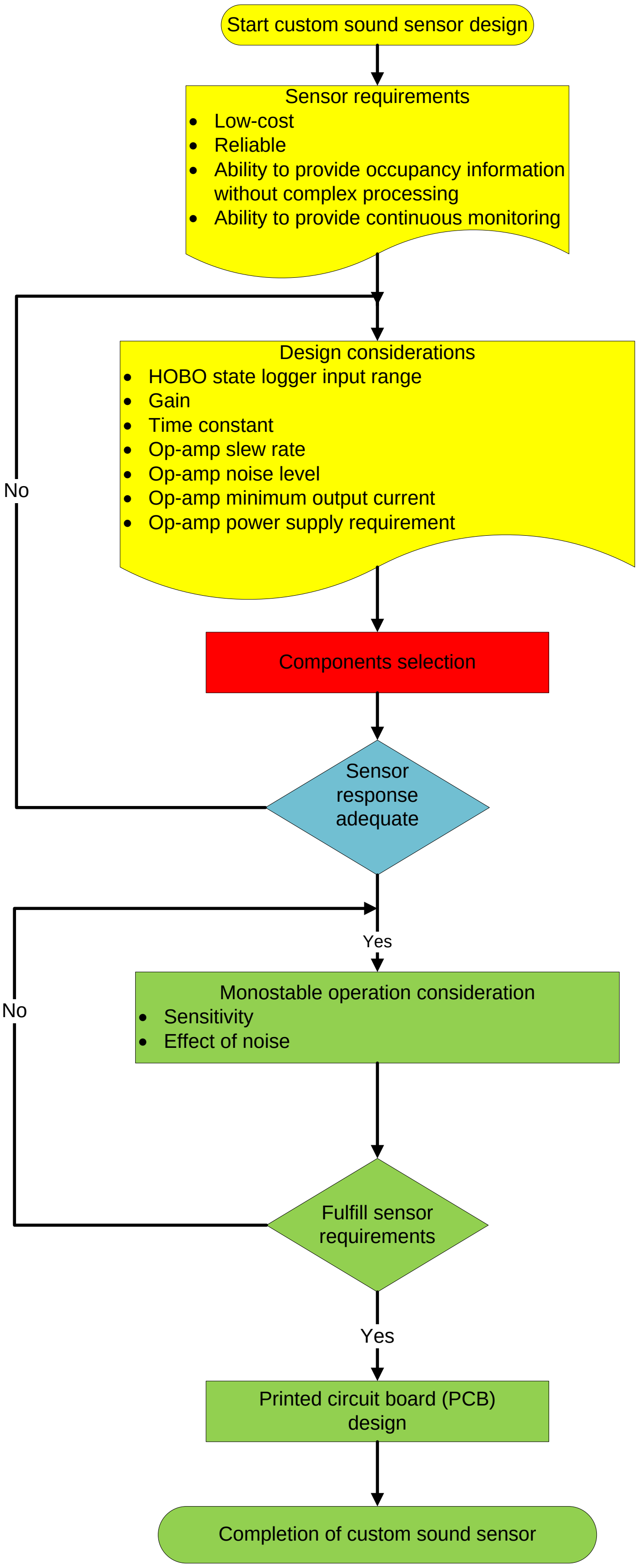


Figure (3.33): Design process of custom sound sensor

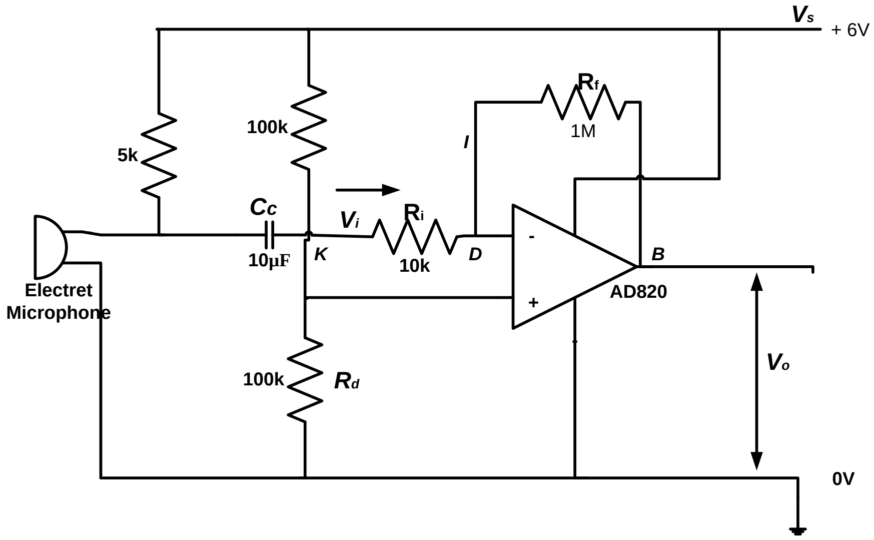


Figure (3.34): Sound sensor circuit –Part A

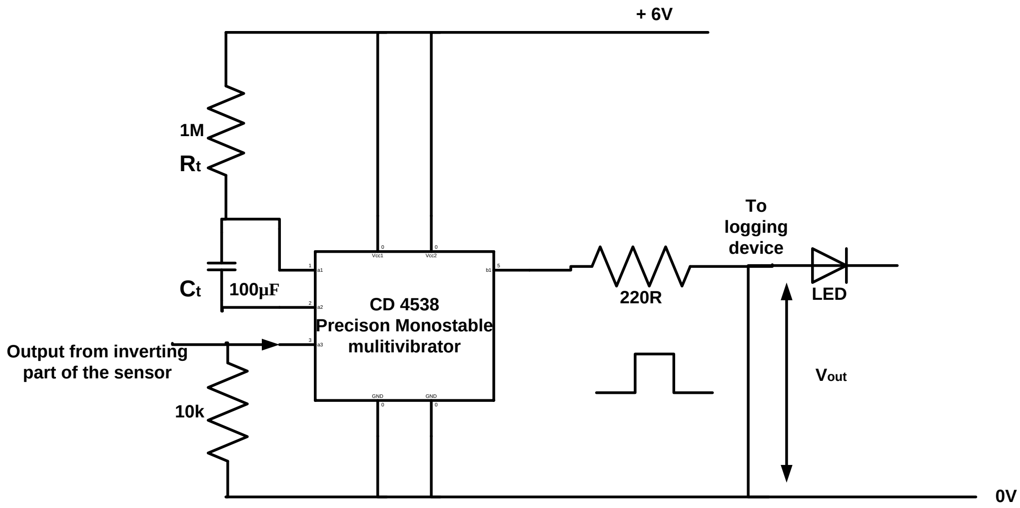


Figure (3.35): Sound sensor circuit –Part B

It was intended for this process to be long enough, so as to trigger the input of the HOBO state logger. A time constant of 100s was selected for sensor development.

$$T = R_t C_t \tag{3.2}$$

Where *R* is resistance and *C* is capacitance

The Electret microphone sensitivity is 0dB=1V/μbar at 1 kHz, converting this parameter to voltage (20mV) informed the amplification factor used, such that microphone input voltage was not amplified above HOBO sensor input range.

Hence, a gain of 100 was selected for the design. Sizes for the input resistor R_i and feedback resistor R_f were chosen as 10k and 1MΩ respectively. It should be noted that what is crucial for gain selection is the resistor ratio, rather than the actual resistance value. However, it is important to use a high value for R_i, due to its high input impedance at the inverting terminal of the Op-amp, *D*.

Table (3.8): Selected AD820 characteristics

Single power supply	5V to 36V
Minimum Output current	15mA
Slew rate	3V/µs
Low noise	16nV/$\sqrt{Hz}$@1kHz

From table (3.8), minimum output current = 15mA, hence *V/I*= 400Ω is the minimum load that can be fed to the op-amp. A design value of 10kΩ was used; this is considered sufficiently high such that it does not load the microphone. Slew rate (SR) is the maximum rate of change of a signal at any point in a circuit, and it constitutes one practical limitation of an op-amp. For a waveform not to be affected by SR limitation, its value must satisfy the condition below:

$$\mathrm{SR} \geq 2\pi f\, V \tag{3.3}$$

Where f is the operating frequency and V is the peak-peak voltage of the waveform.

The maximum signal SR in the experiment was 0.04V/µs. SR for AD820 at 3V/µs is considered adequate. The non-inverting terminal was connected to ground via a resistor, R_d= 100k and this value was carefully chosen, such that both terminals (inverting and non-inverting), will have roughly the same resistance to ground. Otherwise, the two input bias currents can create different voltages at the inputs, which can produce steady d.c at the output, *B*, (even when there is no external input voltage). This is undesirable for the sensor operation. A coupling capacitor, C_c, is added to block any unwanted direct current (D.C) that may overload the amplifier. The need to extract occupancy information from sound events, as regards, presence status: occupied and unoccupied, without any complex data processing prompted the use of a mono-stable IC component. Initially, a 555 timer IC implemented as a

Schimdtt trigger, was coupled to the Op-amp's output, *B*. However, this configuration was not always sensitive to sound events. The 555 timer was also applied as a mono-stable operation, without any significant improvement to sensor's sensitivity. This problem was resolved when a CD4538 precision mono-stable IC was implemented in the design.

3.5.2 Principle of operation

When sound enters the microphone, it produces around 20mV which is then amplified 100 times producing 2V at the output of the Op-amp. The voltage divider in the circuit splits the supply voltage equally at point *K*. The output voltage from the op-amp changes the state of the mono-stable vibrator (CD4538), staying high, indicating occupancy presence, if its input voltage is 2/3 of the voltage at point *K*, or otherwise when this voltage is less than this condition, indicating vacancy. This voltage typically about 2V, charges the 100μF capacitor, and discharges through the *1M*Ω resistor. Signals are then acquired using the HOBO state sensor at one minute intervals. Initial sensor tests showed that the output mostly stayed high. However, when a 22pF de-noising capacitor was attached to the sensor's power supply terminal, its sensitivity was improved. Further test results are shown in figure (3.36) and (3.37).

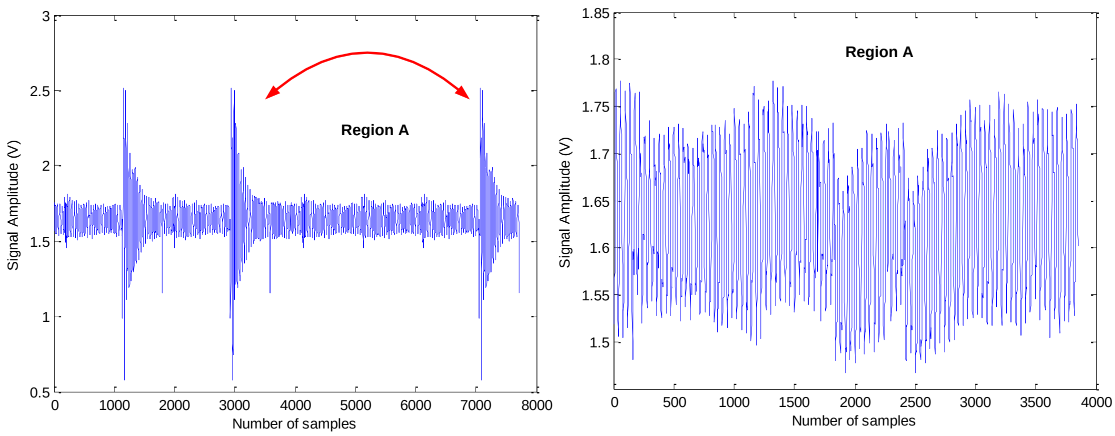


Figure (3.36): Sensor response before mono-stable operation

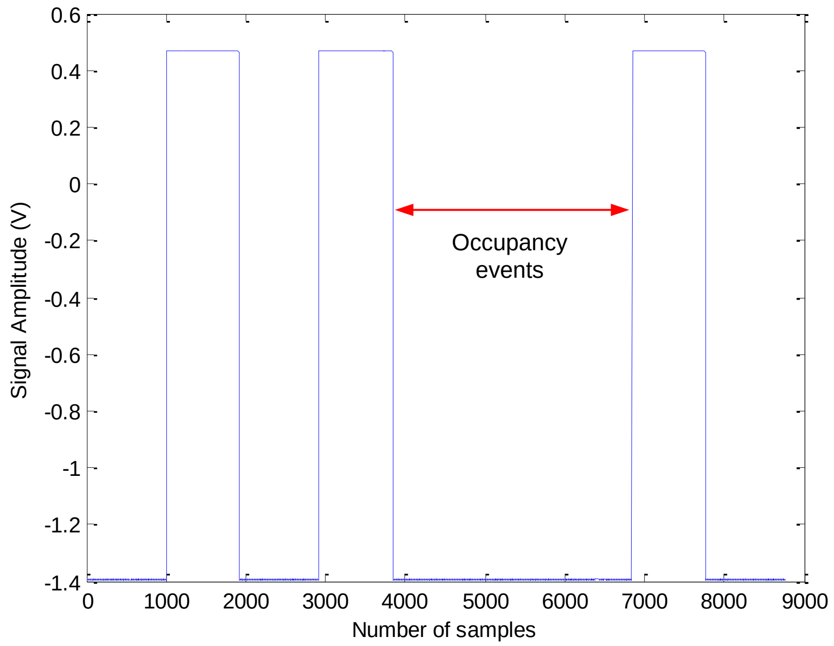


Figure (3.37): Sensor response after mono-stable operation

Using the *LABVIEW* NI USB-6008 data acquisition kit, sound measurements were sampled continuously with 1000 samples collected at a rate of 2 kHz, and subsequently analysed in *MATLAB* environment. Clearly, occupancy events are not obvious before the mono-stable action, as shown in figure (3.36), and may require further complex processing such as FFT analysis to establish occupancy information. However, from figure (3.37), occupancy events are easy to deduce. Figure (3.38) shows the PCB layout for mass producing the custom sound sensor.

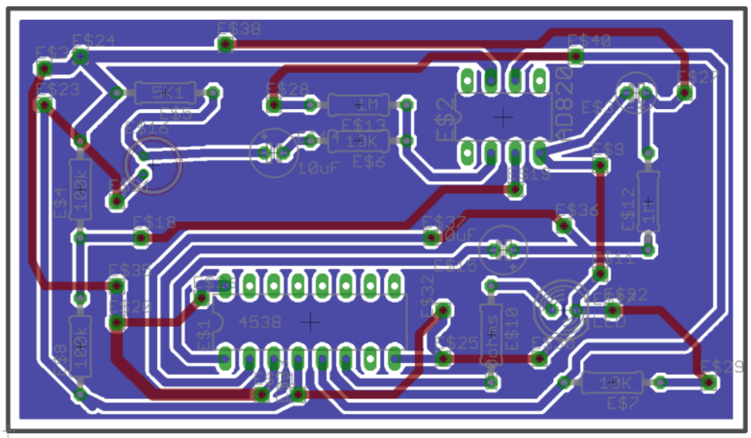


Figure (3.38): PCB layout for the sound sensor

3.6 Occupancy validation method

An infrared camera was mounted in the test area to capture occupants' traffic. Video capture and recording used an ordinary laptop, with images captured at a one-minute interval. This was also found to be a more efficient approach than capturing live streaming video, with no significant loss of resolution. Occupancy numbers validation was carried out by manually counting the number of occupants in the image, see figure (3.39). This information was referred to as actual occupancy numbers in this work, and was used for model training and testing. Image feeds may be corrupted as a result of software updates, or laptop may have needed restarting when image feeds became static, such images were excluded from data analysis.

Figure (3.39): Camera screen shot used for occupancy validation.

3.7 Chapter summary

This chapter has presented a detailed description of an experimental design adopting a multi-sensory instrumentation strategy for occupancy estimation within an open plan office space. EMI mitigation strategy for CO_2 sensors, and details of a custom sound sensor implemented in the multi-sensory strategy adopted in this research were discussed.

In addition, from the studies within this chapter, several points can be drawn as follows:

- Indoor climatic measurements (including case temperature monitoring) are capable of providing occupancy related information.
- The custom sound sensor is able to generate reliable and repeatable occupancy related information without any complex data processing.
- An EMI mitigation strategy proposed in this research provided an insight on the need for good analogue design for building sensors particularly CO_2 sensors.

CHAPTER 4

INTELLIGENT DATA PROCESSING FOR BUILDING OCCUPANCY ESTIMATION

4.0 Introduction

In this chapter, an advanced data processing methodology that employs a variety of intelligent data processing techniques for processing sensor data generated from the instrumentation strategy described in chapter three is presented. Such advanced procedures are needed in a sensor fusion strategy to process raw sensor data in order to arrive at reliable building occupancy estimates. This chapter is organised as follows: section 4.1 describes the data processing methodology. Section 4.2 presents an overview of *WEKA* data mining tool containing some of the algorithms implemented in this research. Section 4.3 previews various tasks (such as handling missing values and time stamp synchronization) to prepare data for further processing. Section 4.4 deals with the extraction of features from pre-processed data, while Section 4.5 presents a detailed account of the feature selection process including feature relevance and redundancy analysis, information theoretical based features ranking, and a correlation based features selection. Section 4.6 summaries the chapter.

4.1 Brief overview of data processing methodology

Estimation of occupancy levels from indoor environmental data using a sensor fusion strategy may require the use of a variety of algorithms. Depending on the system architecture and the goal of the fusion process, different algorithms may be appropriate at various stages of system development (Hall and Llinas, 2001). It becomes time consuming to optimise the performance of these algorithms for different stages. Therefore, it becomes handy to employ any reliable and robust platform (such as *WEKA*) with different machine learning algorithms. In the proposed methodology, the features ranking and selection processes were carried out in *WEKA* (a brief overview of *WEKA* is presented in section 4.2).

Figure (4.1) depicts an overview of the proposed data processing methodology for occupancy levels estimation. Fusion of raw sensor data may not produce reliable occupancy estimates due to the presence of irrelevant and redundant information (Guyon and Elisseeff, 2003), (Kohavi and John, 1997). Hence, it becomes necessary to use relevant features for system development. A feature can be defined as a property describing an instance that may be used to determine its classification (Wilson, 1998). Sensor data were initially read into a *MATLAB* environment, and then synchronised using the same clock. A low pass filter with a cut-off frequency of 2Hz was applied to reduce noise in the data. Once this was done, features of interest from individual sensors were extracted. These features possess different predictive capacities for occupancy estimation. A symmetrical uncertainty measure analysis was applied for determination of the predictive strength of all the features under investigation. Features were then successively passed on to a genetic algorithm based feature selector to search and identify an optimal feature subset.

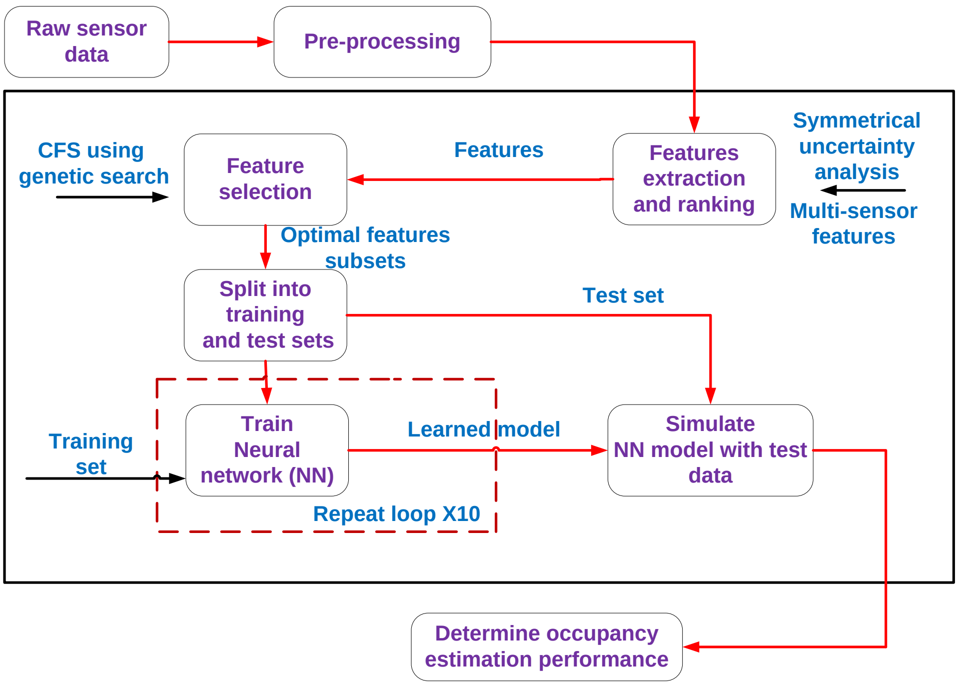


Figure (4.1): The proposed data processing architecture

The candidate or optimal features were those selected the highest number of times throughout the features combination process, appearing 80% of the time. The optimal features subset was split in to training and testing sets, before both were passed on to a neural network model for occupancy levels estimation. The training process was repeated for 10 times to increase the probability of reaching a global solution. Details of the data processing methodology are presented in the subsequent sections.

4.2 An overview of *WEKA* data mining tool

The perceived need for a platform that would afford researchers the opportunity to easily access state-of-the-art machine learning algorithms, saw the introduction of *WEKA* to the machine learning community in 1992 (Hall et al., 2009). More recently, *WEKA* is been applied increasingly in academic fields and commercial settings (Mamidi et al., 2012), (Piatetsky-Shapiro, 2005). This trend could have been made possible owing to the fact that the software is distributed under the GNU General Public License (i.e. open source software). It is also supported by a popular textbook for data mining (Witten and Frank, 2000).

WEKA contains implementation of algorithms for classification, clustering, feature selection, data visualization, data pre-processing and rule association mining. It supports a native file format (ARFF), Excel CSV, *MATLAB* ASCII files etc, and database connectivity through JDBC. Data can be pre-processed using a large number of methods (over 75), including operations such as discretization, handling missing values, normalization etc. *WEKA* contains more than 100 classification algorithms. Classifiers are divided into lazy methods (nearest neighbour and variants); Bayesian methods (Naive Bayes, Bayesian nets, etc.); tree learners (C4.5, Naive Bayes trees, M5); function-based learners (linear regression, SVMs, Gaussian processes); rule-based methods (decision tables, OneR, RIPPER); and miscellaneous methods. Feature selection and classification algorithms such as symmetrical uncertainty analysis, NN, SVM, etc, are useful for development of the occupancy detection system studied in this thesis. Some occupancy detection methodologies in the literature have implemented similar data mining software for system development (Lam et al., 2009b), applying the Accelerated Statistical Learning (ASL) for feature selection. *WEKA* has been used for feature fusion for occupancy detection systems (Mamidi et al., 2012).

4.3 Data pre-processing

The data pre-processing stage aims to prepare raw data collected for feature extraction. It consists of synchronization of the data set in the same time domain, outlier removal, and handling missing values. Figure (4.2) illustrates the sequence of tasks in the data pre-processing stage. Although all data loggers deployed for data collection were synchronised with one clock, individual data loggers' clocks were out of sync with each other by few seconds, usually less than a minute. This was assumed not to have any significant ramification for development of the occupancy detection system, as indoor environmental parameters rarely vary significantly in this time frame. Time-stamp synchronization issues also occurred when there was need to charge batteries used to power CO_2 sensors, or when the laptop used for gathering actual occupancy numbers, required restarting after image feeds from the infrared camera became static. The time –stamp of occupancy profile extracted from such images was usually misaligned with clocks of other data loggers. The synchronisation task was time consuming, as this was done manually. Attempts to automate this process proved difficult, as the time delay for different days and sensors varies, coupled with missing data instances as a result of the instrumentation limitation mentioned earlier in this section. Missing values were removed from data. Once synchronisation was done, outliers were removed using *MATLAB* scripts.

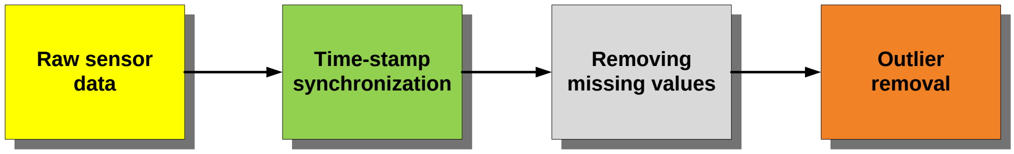


Figure (4.2): Data pre-processing stages

4.4 Features extraction

The nature of features extracted in this study was informed by several studies in the literature review. The choice of features extracted from sensor data play a crucial role in the accuracy of a trained machine learning model. Hence, it is vital that features extracted from sensor data capture as much potential for occupancy information as possible. Different sets of features can be explored for occupancy estimation, and choice of features extracted is dependent on the parameters observed, which may be largely influenced by factors such as available instrumentation and cost. Each monitoring campaign may produce distinct sets of features that capture occupancy related information, since the indoor climatic dynamics of various spaces may be different.

Table (4.1): Mathematical description of features investigated

Air temperature, relative humidity, case temperature, CO_2 and VOC measurements	
First order difference (FDIFF_notation)	raw (i) - raw (i) -1
Second order difference(SDIFF_notation)	raw_FDIFF (i) - raw_FDIFF (i) -1
Time (t) moving average (AVR_notation)	$((\sum_{i-(t-1)}^{i} raw(i))/t(\text{minutes}))$
Approximate area under the curve for between two data instances (A_notation)	$\int_{t(i)-1}^{t(i)} FDIFF(i) \approx FDIFF(i) - FDIFF(i-1)(f(FDIFF(i)) + f(FDIFF(i-1))/2$
Variance (VAR_notation) Where U= Number of sensors, and μ = mean	$\sqrt{\frac{1}{U}\sum_{i}^{N}(AVR)_i}$-$\mu_i$
Sound and PIR data	
Occupied times : Total duration of occupancy as detected by the sound or PIR sensor in a given time interval (TOS_notation)	$\sum_{i}^{U}(AVR\_t(minutes))_i$
Number of pulses : Total number of state changes as detected by the sound or PIR sensor in a given time interval (TONP_notation)	$\sum_{i}^{U}(AVR\_t(minutes))_i$
High to Low : Total number of state changes from high to low as detected by the sound or PIR sensor in a given time interval (THI_LO_notation)	$\sum_{i}^{U}(AVR\_t(minutes))_i$
Low to High : Total number of changes from low to high as detected by the sound or PIR sensor in a given time interval (TLO_HI_notation)	$\sum_{i}^{U}(AVR\_t(minutes))_i$

New sets of features were created based on pre-processed sensor data. The features are intended to capture temporal variations in indoor climatic measurements. Variance has been computed in an attempt to measure the amount of spread of occupancy related events, while the area under the curve was intended to capture the cumulative effect of any indoor variable for a given time interval during occupancy and vacancy periods. Both first and second order differences of monitored indoor climatic variables were intended to measure their associated changes with occupancy entropy in an observed space. Features analysed alongside their description are given in table (4.1).

4.5 Feature selection process

The sensors deployed for monitoring possess different capacities for capturing occupancy information. Hence, the need to develop a systematic selection process to identify the relevant sensors (features). A central objective of this study is to investigate which combination of environmental ambient sensors provides the most relevant information for detection of occupancy numbers. Features obtained from individual sensors in the sensing network were examined to explore their suitability for occupancy estimation using information theory based analysis. In this section, an overview of the feature selection process is presented. The data processing methodology was tested on two different data sets in order to examine its robustness. In this thesis, data set one refers to data obtained from test area two, while data set two was collected from test area three as indicated in table (3.5).

4.5.1 Feature relevance and redundancy

The first action in the feature selection process is determining what sensor features are relevant or redundant for occupancy estimation. A number of definitions exist for feature ‘relevance’ in the machine learning literature, and each depends on the goal of the feature selection task (Blum and Langley, 1997). For the purpose of brevity and appropriateness for the occupancy estimation task, a popular definition given in John et al. (1994) is adopted in this thesis. The definition is stated as follows; Let F be the full set of features, F_i be a feature, S_i = F-$\{f_i\}$, where $S_i = F/F_i$, is the set F with the F_i removed from F. Let C denote the class label. And let P denote the conditional probability of the class label C given a feature set. The statistical relevance of a feature can be formalised as:

Definition 1 (Relevance) *A feature F_i is relevant iff*

$$P\left(C \middle| F_{i,}S_i'\right) \neq P(C|S_i') \tag{4.1}$$

Otherwise, the feature F_i is said to be irrelevant

Definition 2 (Weak relevance) *A feature F_i is redundant iff*

$$P\left(C \middle| F_{i,}S_i\right) = P(C|S_i), \text{ and}$$

$$\exists\ S_i' \underline{C} S_i \text{ , such that } \quad P\left(C \middle| F_{i,}S_i'\right) \neq P(C|S_i') \tag{4.2}$$

Corollary 1 (Irrelevance) *A feature F_i is irrelevant iff*

$$\forall S_i' \underline{C} S_i, P\left(C \middle| F_{i,}S_i'\right) = P(C|S_i') \tag{4.3}$$

The implications of the definitions are described according to Yu et al. (2004);

"*Strong relevance of a feature indicates that the feature is always necessary for an optimal subset; it cannot be removed without affecting the original conditional class distribution. Weak relevance suggests that the feature is not always necessary but may become necessary for an optimal subset at certain conditions. Irrelevance (following Definitions 1 and 2) indicates that the feature is not necessary at all. An optimal feature subset could include all relevant features, no irrelevant features, and a subset of weakly relevant features*"

Zhao et al. (2010) also asserted that;

"*Definition 1 suggests that a feature can be statistically relevant due to two reasons: (1) it is strongly correlated with the class; (2) it forms a feature subset with other features and the subset is strongly correlated with the class*".

Furthermore, irrelevant features can be removed from a candidate optimal feature subset without any effect on a classifier (machine learning model) performance. Definition 2 is not clear on how to distinguish between important relevant features and unimportant weak features (Okun, 2011). This introduces the concept of feature redundancy, which can be examined in terms of correlation. Two features can become redundant if their correlation values are similar (i.e. they are completely correlated).

4.5.2 Correlation measure based feature relevance analysis

Correlation measure is widely used for feature relevance analysis, and the common types applied to analyse any relationship between random variables are; Linear and the non-linear correlation measures. Linear correlation analysis uses the well-known known *Linear or Pearson correlation coefficient* (ρ). For a given pair of variables *(X, Y)*, ρ is given by equation 4.4.

$$\rho = \frac{\sum_i (x_i - \overline{x_i})(y_i - \overline{y_i})}{\sqrt{\sum_i (x_i - \overline{x_i})^2}\ \sqrt{\sum_i (y_i - \overline{y_i})^2}} \tag{4.4}$$

Where $\overline{x_i}$ is the mean of *X*, and $\overline{y_i}$ is the mean of *Y*. The value of ρ lies between -1 and 1, if *X* and *Y* are completely linearly dependent (correlated). If *X* and *Y* are completely independent (uncorrelated), ρ = zero. Linear correlation may be useful for obtaining a quick indication of relevant feature subsets; it helps to remove features with near zero correlation with the class. However, it is not always safe to assume linearity for real-world features, besides a linear correlation measure may not be able to capture non-linear correlations among features (Yu et al., 2004). In this research, in order to ensure a robust feature relevance analysis in the feature selection process, an information-theoretical based concept of entropy was applied. This is a widely used non-linear correlation measure and has also been applied for occupancy detection (Lam et al., 2009b).

4.5.3 Information theory - Symmetrical uncertainty analysis based feature ranking

Sensing domain in this thesis refers to any measurement or feature derived from a specific physical variable in the observed environment such as CO_2, sound etc. The purpose of the feature ranking for each of the sensing domain is to evaluate the degree of association between each feature and the number of occupants.

In information theory, entropy is a measure of the amount of uncertainty of a particular random variable. The entropy of a random variable *Y* with a probability mass function is defined as in equation (4.5). For a pair of random variables (X, Y), if the outcomes of X are known, then the amount of uncertainty for realisation of variable Y is given by the conditional entropy $H(Y|X)$, which is defined by equation (4.6).

$$H(Y) = -\sum_i p(y_i) \log_2 p(y_i) \tag{4.5}$$

$$H(Y|X) = -\sum_j p(x_j) \sum_i p(y_i|x_j) \log_2 p(y_i|x_j) \tag{4.6}$$

Where $p(y_i)$ is the prior probabilities for all values of Y and $p(y_i|x_j)$ is the conditional probability of y_i given x_j.

Suppose Y is treated as different classes in the actual occupancy data, and X as the features extracted from various indoor environmental sensors. $H(Y|X) = 0$, if all feature subsets belongs to the same class, indicating that there is no uncertainty between both variables. While, $H(Y|X) = 1$, if a feature subset is totally random to a class. The amount by which the entropy of Y decreases indicates the additional information about Y provided by X, and is called information gain as shown in equation (4.7). Information gain measures the dependence or common uncertainty between a feature and actual occupancy data, and it is defined as:

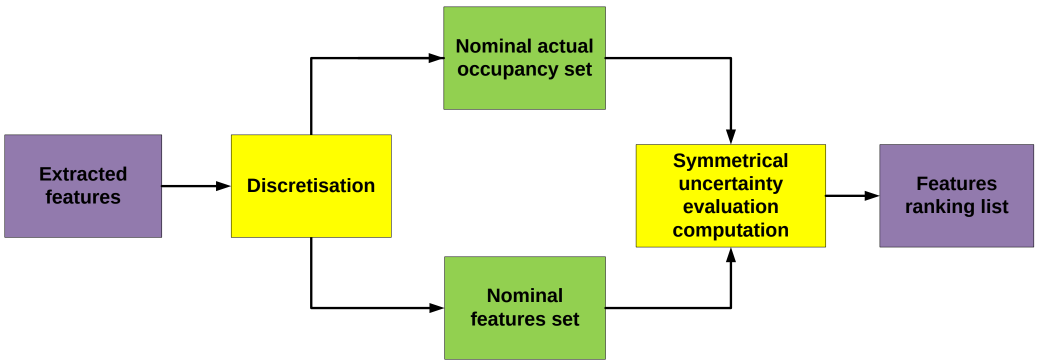
$$I(Y, X) = H(Y) - H(Y|X) \tag{4.7}$$

The mutual information gain between two random variables X and Y is symmetric, and this is a desirable property when measuring correlations between features. However, it is biased for features with more values. Therefore, in this thesis, symmetric uncertainty (SU) is employed to determine the predictive strength of features investigated for occupancy numbers estimation. SU compensates for information gain's bias towards features with more values, and normalizes its values from 0 to 1 to ensure they are comparable. It also treats pairs of features symmetrically, and averages the values of two random variables; and therefore does not have any bias problem. This methodology has been proven to be efficient in removing redundant features in many machine learning applications, thus improving a classifier's accuracy (Hall, 1999), (Senthamarai Kannan, 2010). Features can be ranked according to their predictive abilities. If the symmetric uncertainty evaluation measure of a feature to the actual occupancy data is low, it implies that the feature has poor predictive ability for occupancy levels, and vice-versa.

$$SU(Y,X) = 2\left[\frac{I(Y,X)}{H(X)+H(Y)}\right] \tag{4.8}$$

4.5.4 Feature ranking –data set one

Extracted features and actual occupancy data were discretised into nominal states. This was carried out in order to have a common basis for the evaluation of SU measure values for the different types of features (Okun, 2011). SU value of each sensor feature was obtained based on equation (4.8), and after which a rank of the features set investigated was obtained. The Figure (4.3) shows the data flow in the feature ranking process.



Figures (4.3): Data flow in the feature ranking process

Redundant features in this study are considered as features with the same predictive strength, as per symmetrical uncertainty evaluation measure value. These features have been excluded from the feature selection analysis. Such an action has been known to improve the performance of a machine learning model (John et al., 1994), (Ding and Peng, 2003). Features with SU values in red were excluded from the further analysis in the features selection process, as shown in the features ranking tables (4.2), (4.3) and (4.4). The results in the tables show the average SU values of features investigated.

Table (4.2): Features ranking- CO_2, Relative humidity, ambient temperature and case temperature measurements (data set one) with excluded features (in red).

	AVR	FDIFF	SDIFF	AF_DIFF	AS_DIFF	VAR
CO_2						
SU value	0.273	0.280	0.174	0.275	0.216	0.095
CAS						
SU value	0.203	0.135	0.093	0.141	0.103	0.216
TEMP						
SU value	0.148	0.138	0.065	0.143	0.060	0.053
RH						
SU value	0.105	0.086	0.074	0.086	0.079	0.053

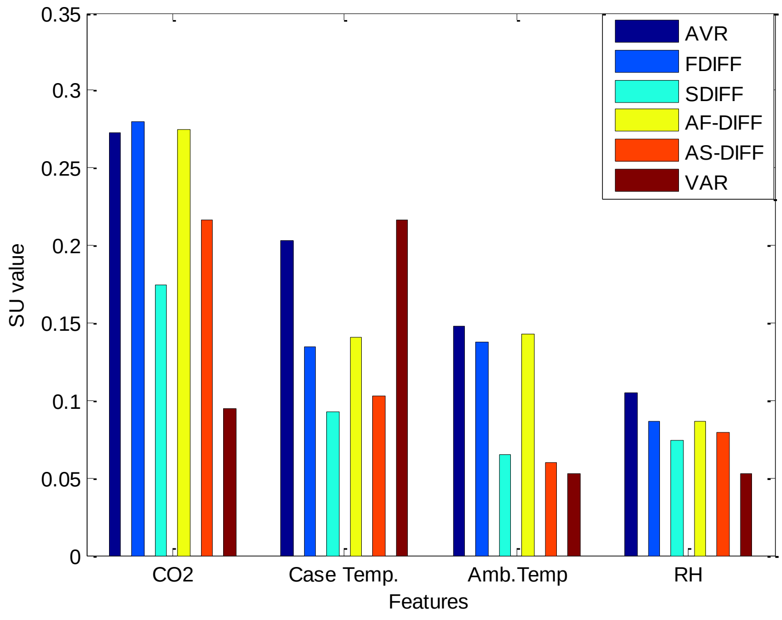
Table (4.3): Features ranking - PIR sensor measurements (data set one) with excluded features (in red).

	TOTAL	FDIFF	SDIFF	AF_DIFF	AS_DIFF	VAR
TOS						
SU value	0.327	0.138	0.112	0.112	0.100	0.293
TONP						
SU value	0.323	0.138	0.118	0.160	0.125	0.289
THI_LO						
SU value	0.323	0.155	0.109	0.133	0.120	0.289
TLO_HI						
SU value	0.323	0.133	0.094	0.130	0.120	0.289

Table (4.4): Features ranking -Sound measurements (data set one) with excluded features (in red).

	TOTAL	FDIFF	SDIFF	AF_DIFF	AS_DIFF	VAR
TOS						
SU value	0.385	0.164	0.130	0.187	0.151	0.312
TONP						
SU value	0.369	0.108	0.112	0.136	0.100	0.367
THI_LO						
SU value	0.370	0.122	0.118	0.154	0.092	0.305
TLO_HI						
SU value	0.370	0.136	0.102	0.136	0.116	0.312

From figure (4.4), FDIFF and AF_DIFF features tend to show stronger predictive power than SDIFF and AS_DIFF features. CO_2 features produced the best correlation with occupancy numbers compared to others. The predictive strength of CO_2 features ranges between 0.280 achieved by FDIFF_CO_2, and 0.095 achieved by VAR_CO_2. The low predictive strength of VAR_CO_2 may suggest that spread of CO_2 levels may be dependent on factors other than occupancy numbers, i.e. space volume and air movement, with the caveat that these results are limited to the particular space under test. AVR_CAS and VAR_CAS showed good predictive ability, with SU values 0.203 and 0.216 respectively. However, the usefulness of these features for occupancy numbers estimation in the observed space is investigated in chapter five. Relative humidity and ambient temperature features show poor correlation with occupancy based on SU analysis, with their best predictive features values reaching 0.148, which is relatively lower compared to features obtained from case temperature and CO_2 measurements. This is not surprising, as it only supports previous research, suggesting both parameters are dominated by space heating and cooling (Dong et al., 2010).



Figures (4.4): CO_2, case temperature, ambient temperature and relative humidity sensors feature ranking - dataset one

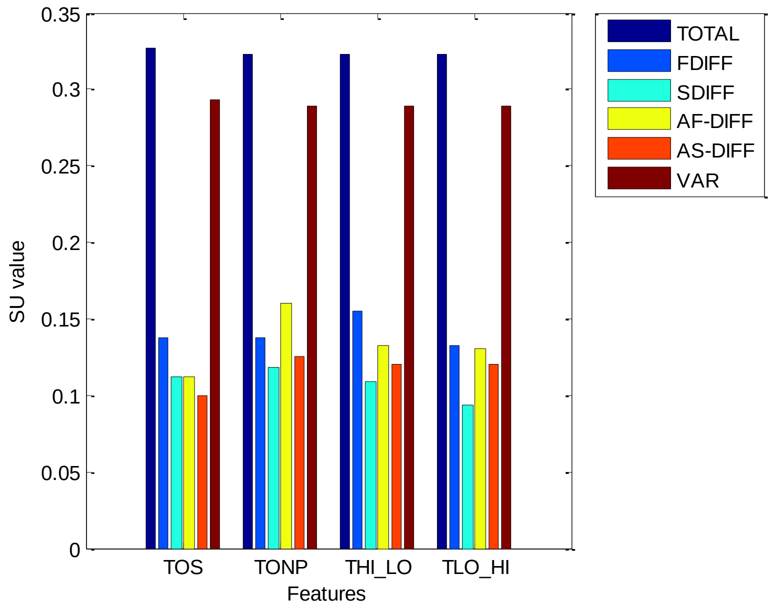
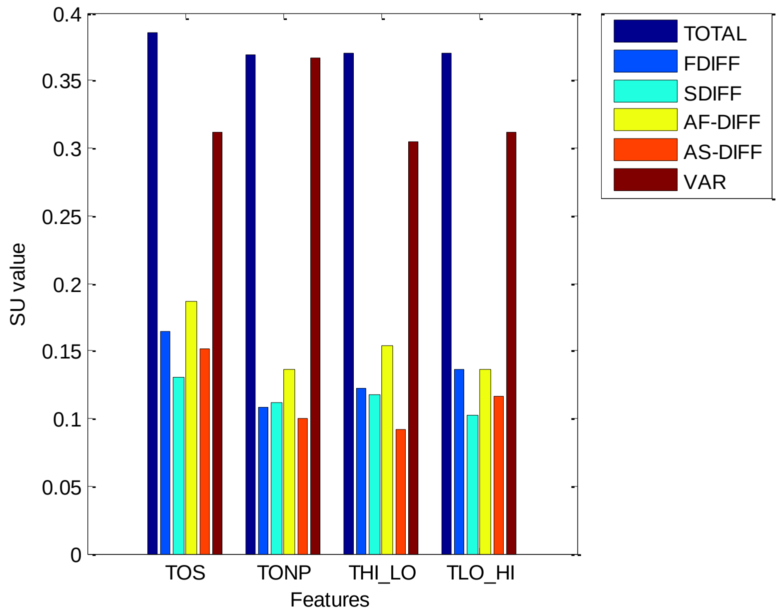


Figure (4.5): PIR sensors feature ranking - data set one



Figures (4.6): Sound sensors feature ranking – data set one

From figure (4.5) and (4.6), the trends are similar for features obtained from PIR and sound measurements. Features such as TOS and VAR, which best capture spatio-temporal changes in the observed environment had the hightest SU values, while others such as FDIFF and SDIFF which monitor paramter entropy with occupancy changes, as well as AF_DIFF and AS_DIFF, were not as good. TOS_SND and TOS_PIR features showed the strongest predictive ability among the individual sensing domain investigated, with SU values for TOS reaching 0.385 and 0.327 respectively.

4.5.5 Feature ranking – data set two

From figure (4.7), VOC features had poor predictive ability for occupancy numbers, with a maximum value of 0.065 for the features investigated. Again, this lends credence to previous research that occupants are not necessarily the major source of VOCs in buildings (EPA, 1991). CO_2 levels have a strong correlation with occupancy numbers, as captured by features such as FDIFF_CO_2 and AF_DIFF_CO_2, with SU values of 0.306 and 0.300 respectively. These values are higher than those obtained from the previous analysis (data set one), although with a caveat that data were collected at different times from the two spaces under test. This trend is expected as the current space (from which data set two was obtained), has a smaller space volume and smaller windows, and does not have ventilation stacks, all of which may impact on the CO_2 levels in the space. All desktop computers in the test area were instrumented for case temperature measurements, and indeed captured sufficient occupancy information to make some of its features good predictors of occupancy, as shown by AVR_CAS and VAR_CAS, with SU values reaching 0.330 and 0.218 respectively. Relative humidity and ambient temperature measurements have been excluded from the feature relevance analysis, due to findings from initial analysis and previous research as mentioned earlier in section (4.5.4).

Table (4.5): CO_2, VOC and case temperature measurements (data set two) with excluded features (in red)

	AVR	FDIFF	SDIFF	AF_DIFF	AS_DIFF	VAR
CO_2						
SU value	0.161	0.306	0.029	0.300	0.059	0.125
CAS						
SU value	0.330	0.167	0.104	0.169	0.112	0.218
VOC						
SU value	0.065	0.030	0.010	0.030	0.015	0.015

Table (4.6): PIR sensor measurements (data set two) with excluded features (in red)

	TOTAL	FDIFF	SDIFF	AF_DIFF	AS_DIFF	VAR
TOS						
SU value	0.171	0.093	0.033	0.109	0.082	0.204
TONP						
SU value	0.208	0.126	0.048	0.123	0.101	0.240
THI_LO						
SU value	0.209	0.126	0.089	0.123	0.097	0.240
TLO_HI						
SU value	0.209	0.126	0.093	0.126	0.105	0.240

Table (4.7): Sound measurements (data set two) with excluded features (in red)

	TOTAL	FDIFF	SDIFF	AF_DIFF	AS_DIFF	VAR
TOS						
SU value	0.320	0.136	0.089	0.131	0.111	0.339
TONP						
SU value	0.281	0.124	0.055	0.136	0.084	0.315
THI_LO						
SU value	0.291	0.111	0.070	0.111	0.075	0.310
TLO_HI						
SU value	0.297	0.124	0.049	0.127	0.084	0.294

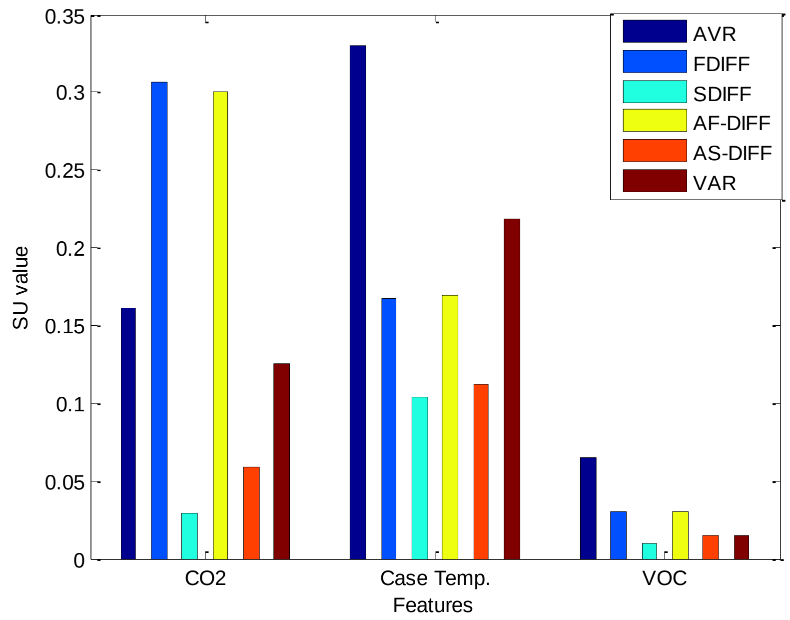


Figure (4.7): CO_2, case temperature and VOC sensors feature ranking –data set two

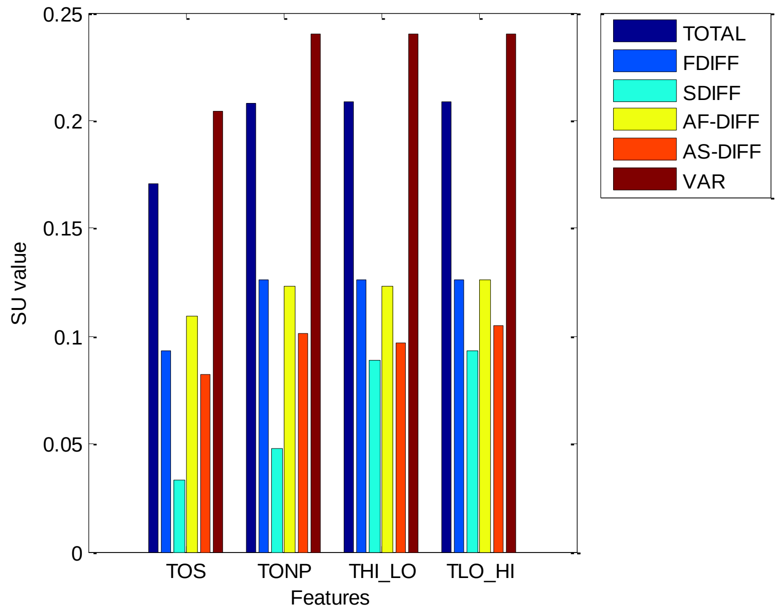


Figure (4.8): PIR features ranking – data set two

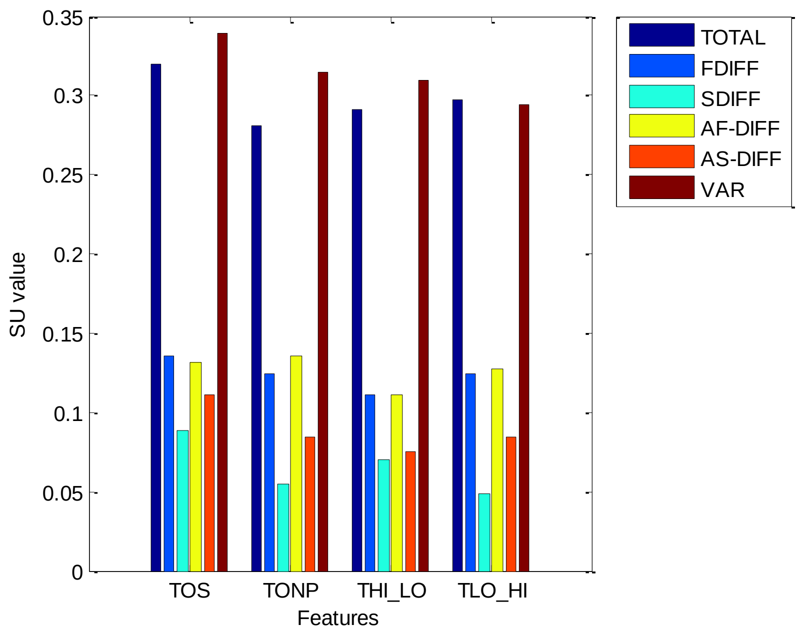


Figure (4.9): Sound features ranking – data set two

VAR features had the highest SU values for PIR and sound measurements, see figure (4.8) and (4.9). The predictive performance of PIR features obtained from test area three with the best SU value at 0.240 achieved by THI_LO_PIR was lower than those recorded in test area two. This may be due to the spatial arrangement of the sensors, as these were positioned close to existing BEMS PIR sensors, which may not capture the movement of all six occupants at all times. Sound features were effective for occupancy monitoring, the highest SU values reaching 0.339 achieved by VOS_SND. This suggests that sound sensing may be useful for capturing the spread of occupancy information in an observed space.

4.5.6 Correlation-based feature selection

Correlation feature selector (CFS) is an algorithm that selects feature subsets according to a correlation based heuristic evaluation function. CFS is biased toward subsets that have high correlation with the class, and uncorrelated with each other (Hall, 1999), hence removing irrelevant and redundant features. Features from each individual sensing domain investigated were passed on to a genetic based CFS filter, to determine an optimal combination of features for occupancy detection. This is because SU and most feature weighting/ranking algorithms are incapable of removing redundant features, since redundant features are likely to have similar rankings (Yu et al., 2004). CFS uses a correlation based heuristics called “merit" to evaluate the worth of features.

$$Merit_S \frac{k\overline{r_{cf}}}{\sqrt{k+k(k-1)\overline{r_{ff}}}} \tag{4.9}$$

$Merit_S$ is the heuristic "merit" of a feature subset, S, containing k features, and $\overline{r_{cf}} = \sum_{f_i} \epsilon S \frac{1}{k} \sum(f_i, C)$ is the mean feature class correlation and $\overline{r_{ff}}$ is the average feature inter-correlation. CFS computes the correlation between feature-feature and feature-class dependence using equation (4.8), and then explores the feature space for an optimal combination of a feature subset with the highest $Merit_S$ value using a heuristic search strategy, with a stopping criterion set at when this value does not increase or after a specific number of iterations. The feature selection process was carried out in *WEKA* (Hall et al., 2009). A detailed description of CFS algorithm can be seen in (Hall, 1999). This algorithm selects the maximum relevant features, prevents the re-introduction of redundant features in the search space, and also works well for small data sets (Zhao et al., 2010). It can identify relevant features where moderate feature dependencies exist (Hall, 1999). However, when features depend strongly on others given the class, CFS can fail to select all the relevant features (Hall and Holmes, 2003). To address this, features with same SU values have been excluded from the feature selection analysis.

4.5.7 Genetic algorithm implemented for feature selection

GAs are increasingly been applied in several engineering studies (Wright et al., 2002), (Guillemin and Morel, 2001), they offer some unique advantages compared to conventional optimisation algorithms. For instance, GAs can use a global search to ensure an optimal or a near optimal solution as opposed to other algorithms where a global optimal solution may not always be guaranteed (Krarti, 2003). In addition, they do not make use of derivatives during their computation, and therefore, optimization of any non-smooth objective function is possible (Krarti, 2003). The feature selection process is seen as a complicated non-linear search problem, where each state in the features space specifies a distinct subset of possible features (Blum and Langley, 1997). It becomes necessary to make use of a robust optimization algorithm (such as a GA) to search for a possible optimal feature subset for occupancy levels estimation. To ensure robustness of the feature selection process, two scenarios are tested, both implementing a GA in searching for an optimal feature subset. Figure (4.10) shows the stages in the feature selection process.

- **Initial population**

An initial population of individuals (a feature subset) was selected randomly, and encoded as chromosomes. These are represented as binary strings within the GA system, as implemented in Holland (1975). An illustration of a typical feature subset containing three features is presented below;

Chromosome 1 (Feature 1): 10010101110101001

Chromosome 2 (Feature 2): 10011101101110111

Chromosome 3 (Feature 3): 11101110010100001

Although, it cannot be guaranteed that these chromosomes are near-optimal or could lie in a local minimum of the solution search space. The first generation subset of individuals is made up of chromosomes from features investigated. The GA system was evaluated for a small population size of 30, so as to provide fast evaluation of each generation, while also ensuring adequate availability of individuals, to maintain variety.

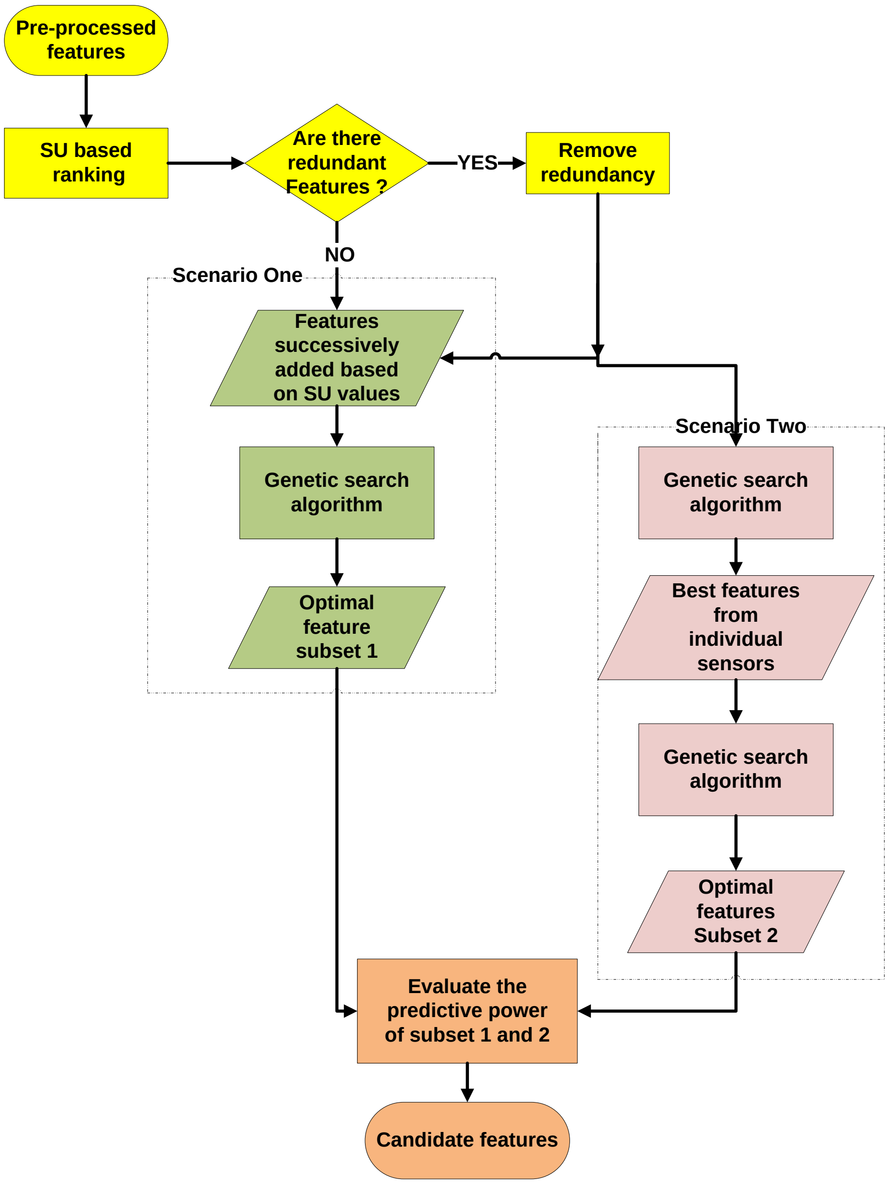


Figure (4.10): Stages in the feature selection process

- **Evaluation function**

The evaluation function is the index used for measuring the fitness of individual feature subsets in the search space. This function is the merit, as defined in equation (4.9). Increasing merits denote increase in the predictive power of feature subsets.

- **Selection**

Selection of individuals in to the next generation was achieved using a roulette wheel selection as in (Goldberg, 1989). This is an important step in the GA operation. Individuals were chosen based on their fitness. It does not always guarantee that the fittest member will proceed to the next generation, as such it allows for even the weakest individual to be selected. However, fitter individuals have the highest probability of being selected.

- **Genetic operators**

Individuals chosen for the next generation are copied or changed using crossover or mutation operators. For the crossover operator, two individuals are chosen to form a new offspring chromosome; it combines different features from a pair of subsets into a new subset. Mutation indicates the probability of a new offspring changing at each locus (position in chromosome). Mutation is responsible for the random addition or deletion of features in a subset. De Jong (1975), who studied GAs for function optimization in a series of parametric experiments, suggested "*that good GA performance requires the choice of a high crossover probability, a low mutation probability (inversely proportional to the population size), and a moderate population size*". Based on the above principle, the following values stated in table (4.8) have been adopted for the GA parameters.

Table (4.8): Values for GA parameters implemented in the n-Rule combination process

GA parameters	
Mutation	0.0333
Crossover	0.6
Populations size	30

- **Termination criteria**

Termination is the criterion a GA uses to stop searching the feature space, and provide the final solution. Many criteria exist for termination of the GA execution once a condition is met, e.g. number of generations, evolution or computation time, fitness threshold etc. Here, a predefined number of generations was used as the stopping criterion for the GA based feature selection analysis.

4.5.8 Scenario one-data set one

Based on the SU analysis, see table (4.2), (4.3), and (4.4), features from each individual sensing domain were passed on to a genetic based filter, starting with the features with highest predictive strength, successively down to the least.

Let n Rule represent a combination that has n features in it. The $Merit_S$ score for each multi-sensory combination was computed using CFS, where n-Rule =1, 2, 3…6. Table (4.9), shows the features selected after each successive combination. A 10-fold cross validation process was employed throughout the n rule combination process, such that for a feature to be selected, it had to be chosen once in each of the 10 runs for each n-Rule multisensory features combination. In this research, the most effective features were evaluated as those selected the highest number of times throughout the n rule combination process, appearing over 80% (or five times of the six) in the process. These were considered as the most dominant features, and may be the best set of features for occupancy number detection within the observed environment. They form the inputs for a machine learning model used for actual fusion process.

From figure (4.11), it is clear the merit score increases with each successive n-Rule combination, indicating increasing predictive ability of selected feature subsets. For n=1, 4 features were selected with merit score of 0.450, increasing to 0.477 with 5 features selected at n=2. This trend continues at n=3 and n=4, with merit score of 0.488 and 0.496 with 7 and 9 features selected respectively. However, the merit score increased marginally after a combination of 24 features (when n=4), by 0.001, and it increases again after 36 features (n=6) were combined. This may be due to over-fitting of the data, as the number of features selected decreased between n=4 and n=5.

The fact that sound and PIR features had the highest SU values (see table 4.3 and 4.4), hence the better predictive power than other features, provides an indication for their dominance in the n-Rule combination process, as shown in table (4.9). This is not surprising, as the other features with good occupancy predictive strength such as case temperature and CO_2 features (even if occupants make up a significant CO_2 source in buildings) may not necessarily track occupancy numbers in the space. The large volume of the open plan space coupled with air infiltration from large windows and ventilation vents may have affected the ability of CO_2 features to track well with occupancy numbers, thereby impacting on their dominance in the feature selection process. The dominance of case temperature features in the selection process may have been weakened by the fact that not all occupants in the space make use of desktop computers, which were instrumented for case temperature measurements. Some make use of their laptops, so case temperature monitoring is unable to detect their presence. From table (4.9), TOS_SND, TOS_PIR, and THI_LO_SND satisfy the stipulated condition used in the analysis, hence they are considered as candidate features for occupancy estimation in the area under test.

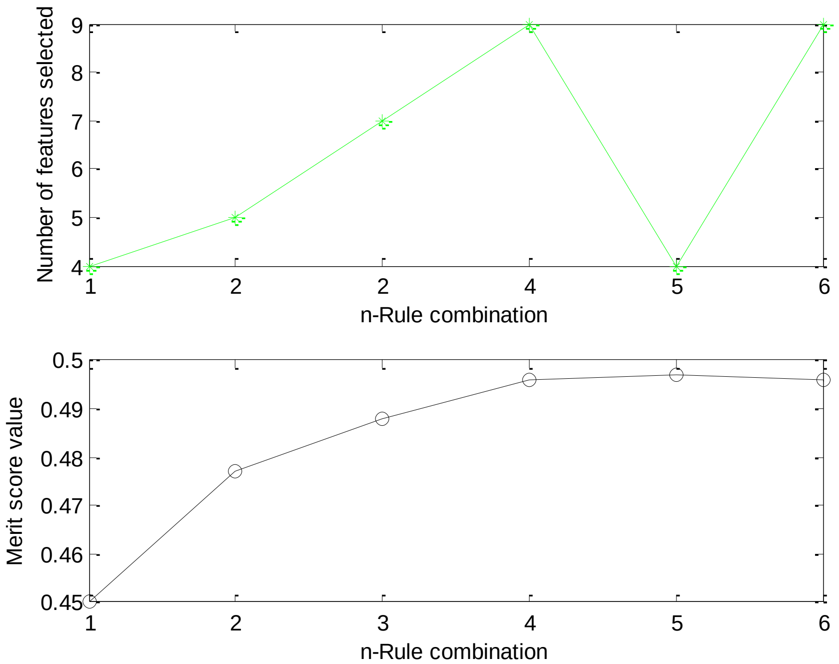


Figure (4.11): Merit score and different n-Rule feature combination (Scenario one – data set one)

4.5.9 Scenario two- data set one

In order to test robustness of the selection process, a different approach from that in scenario one was employed to determine the dominant features. For each individual sensing domain, features were combined to evaluate the best predictive strength of feature subsets for a specific sensing domain, refer to figure (4.10). This analysis was meant to establish the maximum number of feature combinations for an individual sensing domain till the predictive ability (merit score) of a features subset does not increase. The n-Rule combination process applied in scenario one was also used for this.

From table (4.10) and figure (4.12), AVR_CO_2, FDIFF_CO_2, AF_DIFF_CO_2, AS_DIFF_CO_2 appeared to be the dominant feature subset in this sensing domain, and also produced the highest merit value for occupancy estimation. Merit score increased with features combination, till it reached 0.367, and remained constant at this value for subsequent combinations. The trend was the same for case temperature, and ambient temperature with highest merit values reaching 0.258 and 0.195 respectively. All six features were selected and produced a merit of 0.148 for RH measurements, table (4.13) and figure (4.12). While sound features produced merit values reaching 0.434, see figure (4.13). Table (4.11), (4.12), and (4.15) shows the dominant features subsets for case temperature, ambient temperature and sound respectively (the ones with the highest merit score). The feature subset comprising TOS_PIR, THI_LO_PIR and VTOS_PIR with a merit score of 0.396 was selected throughout the n-Rule combination process for PIR measurements, see figure (4.13) and table (4.14).

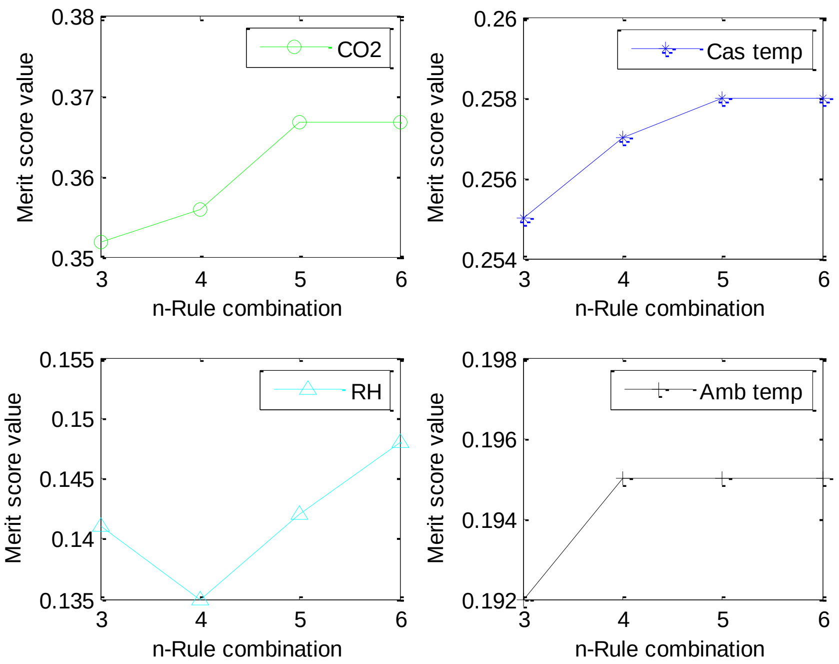


Figure (4.12): n-Rule features combination for individual sensing domain- data set one

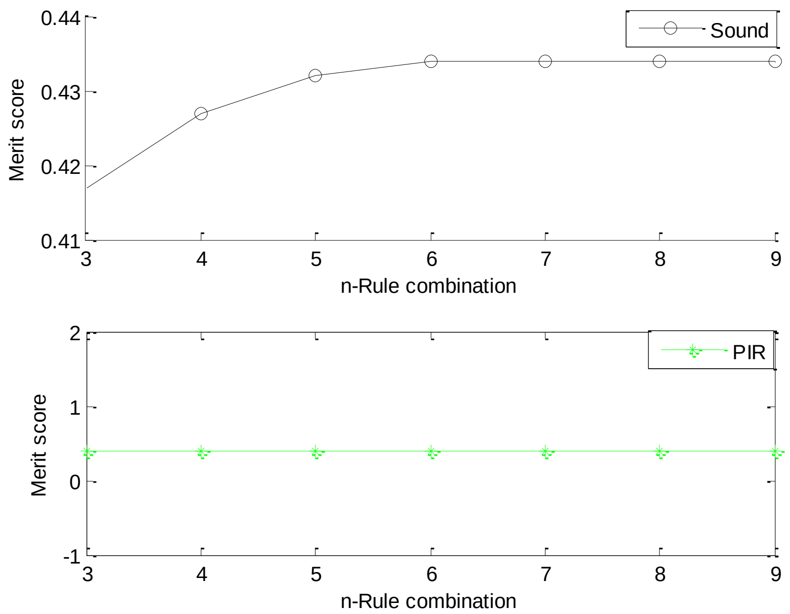


Figure (4.13): n-Rule features combination for PIR and sound sensor –data set one

Table (4.9): Heterogeneous multi-sensory features selection (Scenario one -data set one) indicating the dominant features in colours.

n-Rule combination	Features selected									Merit score
n=1	**TOS_SND**	**TOS_PIR**	**VAR_CAS**	**FDIFF_CO_2**						**0.450**
n=2	**FDIFF_CO_2**	**TOS_PIR**	**TOS_SND**	**THI_LO_SND**	**THI_LO_PIR**					**0.477**
n=3	**TOS_PIR**	**TOS_SND**	**THI_LO_SND**	**THI_LO_PIR**	**VOS_PIR**	**VTONP_SND**	**AVR_CO_2**			**0.488**
n=4	**TOS_PIR**	**TOS_SND**	**THI_LO_SND**	**AVR_CO_2**	**VTOS_PIR**	**VTONP_SND**	**AS_DIFF_CO_2**	**VTLO_HI_PIR**	**VTLO_HI_SND**	**0.496**
n=5	**VAR_CAS**	**TOS_PIR**	**TOS_SND**	**THI_LO_SND**						**0.497**
n=6	**VAR_CAS**	**TOS_PIR**	**TOS_SND**	**THI_LO_PIR**	**THI_LO_SND**	**AVR_CO_2**	**VTONP_SND**	**VTLO_HI_SND**	**VHI_LO_SND**	**0.496**

Scenario two

Table (4.10): CO_2 features combination – data set one

n-Rule combination	**Features selected**				**Merit score**
n=3	AVR_CO_2	FDIFF_CO_2	SDIFF_CO_2		0.352
n=4	AVR_CO_2	FDIFF_CO_2	SDIFF_CO_2	AF_DIFF_CO_2	0.356
n=5	AVR_CO_2	FDIFF_CO_2	AF_DIFF_CO_2	AS_DIFF_CO_2	0.367
n=6	AVR_CO_2	FDIFF_CO_2	AF_DIFF_CO_2	AS_DIFF_CO_2	0.367

Table (4.11): Case temperature features combination – data set one

n-Rule combination	Features selected				Merit score
n=3	AVR_CAS	VAR_CAS	FDIFF_CAS		0.255
n=4	AVR_CAS	VAR_CAS	FDIFF_CAS	SDIFF_CAS	0.257
n=5	AVR_CAS	VAR_CAS	AF_DIFF_CAS		0.258
n=6	AVR_CAS	VAR_CAS	AF_DIFF_CAS		0.258

Table (4.12): Ambient temperature features combination- data set one

n-Rule combination	Features selected		Merit score
n=3	AVR_TEMP	FDIFF_TEMP	0.192
n=4	AVR_TEMP	AF_DIFF_TEMP	0.195
n=5	AVR_TEMP	AF_DIFF_TEMP	0.195
n=6	AVR_TEMP	AF_DIFF_TEMP	0.195

Table (4.13): Relative humidity features combination – data set one

n-Rule combination	Features selected						Merit score
n=3	AVR_RH	FDIFF_RH	SDIFF_RH				0.141
n=4	AVR_RH	FDIFF_RH	SDIFF_RH				0.135
n=5	AVR_RH	FDIFF_RH	SDIFF_RH	AF_DIFF_RH	AS_DIFF_RH		0.142
n=6	AVR_RH	FDIFF_RH	SDIFF_RH	AF_DIFF_RH	AS_DIFF_RH	VAR_RH	0.148

Table (4.14): PIR features combination-data set one

n-Rule combination	Features selected			Merit score
n=3	TOS_PIR	THI_LO_PIR	VTOS_PIR	0.396
n=4	TOS_PIR	THI_LO_PIR	VTOS_PIR	0.396
n=5	TOS_PIR	THI_LO_PIR	VTOS_PIR	0.396
n=6	TOS_PIR	THI_LO_PIR	VTOS_PIR	0.396
n=7	TOS_PIR	THI_LO_PIR	VTOS_PIR	0.396
n=8	TOS_PIR	THI_LO_PIR	VTOS_PIR	0.396
n=9	TOS_PIR	THI_LO_PIR	VTOS_PIR	0.396

Table (4.15): Sound features combination- data set one

n-Rule combination	Features selected						Merit score
n=3	TOS_SND	THI_LO_SND					0.417
n=4	TOS_SND	THI_LO_SND	VTONP_SND	VTLO_HI_SND			0.427
n=5	TOS_SND	THI_LO_SND	VTONP_SND	TLO_HI_SND	THI_LO_SND		0.432
n=6	TOS_SND	THI_LO_SND	VTONP_SND	VTLO_HI_SND	VTHI_LO_SND	AF_DIFF_TOS_SND	0.434
n=7	TOS_SND	THI_LO_SND	VTONP_SND	VTLO_HI_SND	VTHI_LO_SND	AF_DIFF_TOS_SND	0.434
n=8	TOS_SND	THI_LO_SND	VTONP_SND	VTLO_HI_SND	VTHI_LO_SND	AF_DIFF_TOS_SND	0.434
n=9	TOS_SND	THI_LO_SND	VTONP_SND	VTLO_HI_SND	VTHI_LO_SND	AF_DIFF_TOS_SND	0.434

Table (4.16): Features subsets with highest merit score for individual sensing domain- data set one

Individual sensing domain	Features selected						Merit score
CO_2	AVR_CO_2	FDIFF_CO_2	AF_DIFF_CO_2	AS_DIFF_CO_2			0.367
Case temperature	AVR_CAS	VAR_CAS	AF_DIFF_CAS				0.258
Ambient temperature	AVR_TEMP	AF_DIFF_TEMP					0.195
Relative humidity	AVR_RH	FDIFF_RH	SDIFF_RH	AF_DIFF_RH	AS_DIFF_RH	VAR_RH	0.148
PIR	TOS_PIR	THI_LO_PIR	VTOS_PIR				0.396
Sound	TOS_SND	THI_LO_SND	VTONP_SND	VTLO_HI_SND	VTHI_LO_SND	AF_DIFF_TOS_SND	0.434

Table (4.17): Heterogeneous multi-sensory features selection (Scenario two - data set one) indicating the dominant features in colours

n-Rule combination	Features selected								Merit score
n=1	FDIFF_CO$_2$	VAR_CAS	TOS_PIR	TOS_SND					0.450
n=2	TOS_PIR	TOS_SND	THI_LO_PIR	THI_LO_SND					0.473
n=3	TOS_PIR	TOS_SND	THI_LO_PIR	THI_LO_SND	AVR_CO$_2$	VTOS_PIR	VTONP_SND		0.488
n=4	FDIFF_CO$_2$	TOS_PIR	TOS_SND	THI_LO_SND	AVR_CO$_2$	VTOS_PIR	VTONP_SND	AS_DIFF_CO$_2$	VTLO_HI_SND	0.494
n=5	TOS_PIR	TOS_SND	THI_LO_PIR	THI_LO_SND	AVR_CO$_2$	VTOS_PIR	VTONP_SND	AS_DIFF_CO$_2$	VTLO_HI_SND	0.497
n=6	VAR_CAS	TOS_PIR	TOS_SND	VTOS_PIR	VTONP_SND	AS_DIFF_CO$_2$	VTLO_HI_SND	VTHI_LO_SND		0.494

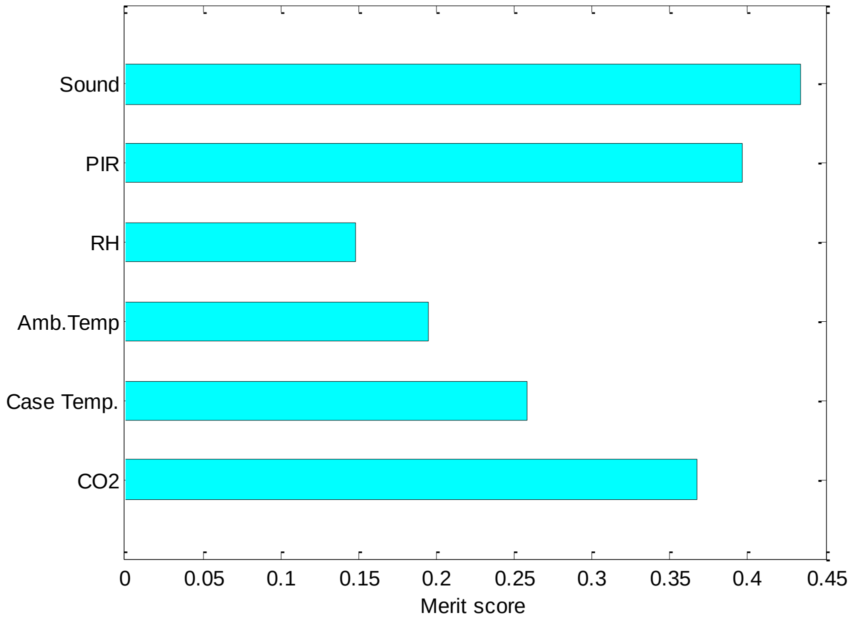


Figure (4.14): Effectiveness of individual sensing domain- data set one

Feature subsets producing the highest merit scores for the individual sensing domain may give an indication of the effectiveness of a particular sensing domain for occupancy estimation. From figure (4.14), sound and PIR sensors appear to produce the most effective sets of features as per the merit score. The strong predictive capacity of sound features may suggest that the custom sound sensor design used is capable of providing reliable information to estimate occupancy numbers in open – plan offices. In order to arrive at an optimal features subset from the different sensor types, only the best feature subsets (which produced the highest merit scores) for each individual sensing domain (refer to table 4.16) were combined in scenario two using a similar n-Rule combination process as in scenario one.

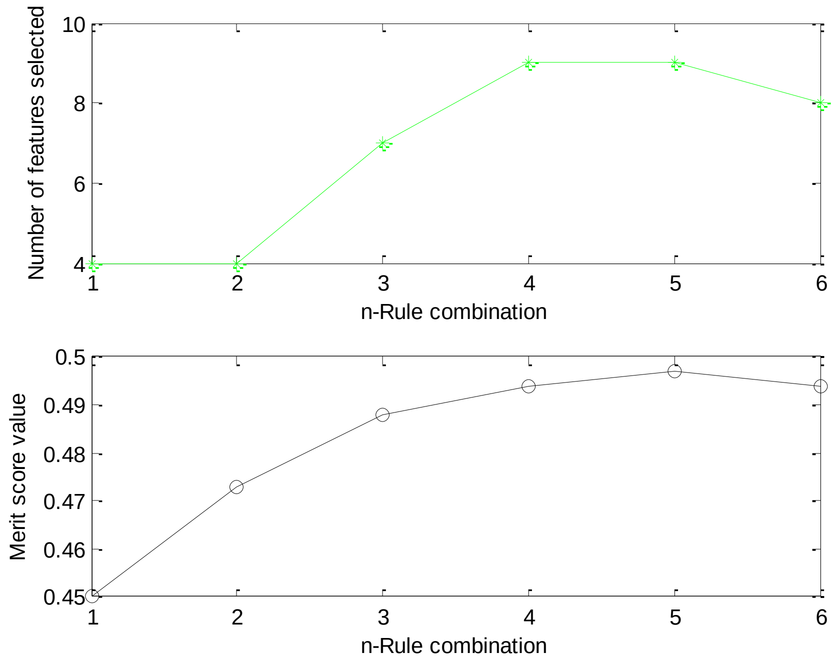


Figure (4.15): Merit score and different n-Rule feature combination

(Scenario one –data set one)

As with the case in scenario one, from figure (4.15) merit score increased with each successive n-Rule combination, till when n= 5 where the number of selected features was constant, although with an increase in the merit score, suggesting over fitting of data. However, both decreased at n=6, suggesting the best merit score for the features subsets may have already been attained. As shown in table (4.17), just two features: TOS_PIR and TOS_SND satisfied the adopted 80% selection criterion in the process, hence both were considered as the dominant features.

A further test which involves evaluating the predictive ability of feature subsets selected from scenario one and two was carried out to arrive at a final feature subset for occupancy estimation. Each feature subset was combined using CFS to ascertain which feature subset produces the highest merit value that can be obtained. The feature subset comprising TOS_PIR, TOS_SND and THI_LO_SND was used as the final fusion input features, having produced a higher merit value than that of scenario one.

4.5.10 Scenario one- data set two

A similar analysis in section (4.5.9) was repeated here,where features from different sensing domain based on SU anlysis presented in tables (4.5), (4.6), and (4.7) were combined. The results also show similar patterns, see figure (4.16). The number of features selected increased with merit score. For n=1, 4 features were selected with merit score of 0.461, increasing to 0.478 with 8 features selected at n=2. This trend continues at n=3 and n=4, with merit score of 0.491 and 0.492 with 10 and 7 features selected respectively. The merit score for n=5 and n= 6 remained constant, as with the number of features. However, number of selected features increasing from 7 to 9 at n=6, may be due to over-fitting.

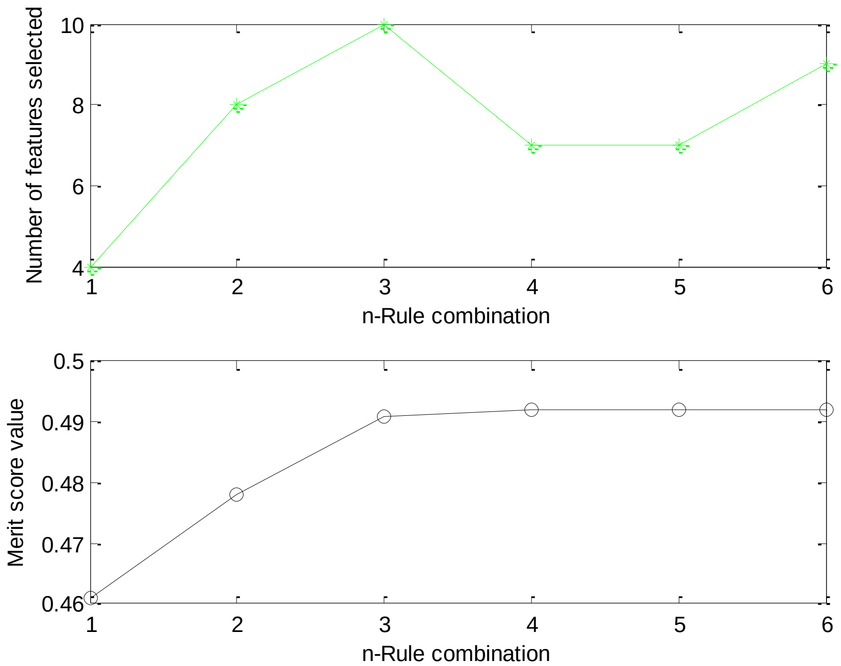


Figure (4.16): Merit score and different n-Rule feature combination

(Scenario one –data set two)

From table (4.18), four features namely: FDIFF_CO_2, AVR_CAS, AF_DIFF_CO_2, and VOS_SND dominated the selection process. This selection supports results of the feature ranking process, where these features indicated high predictive capacity for occupancy estimation. Dominance of FDIFF_CO_2 and AF_DIFF_CO_2 may suggest their effectiveness in capturing temporal changes in CO_2 levels associated with occupancy entropy in the space. Case temperature monitoring is also effective for this environment with the selection of AVR_CAS. In scenario two, further analysis is presented to investigate consistency of the selection process.

4.5.11 Scenario two- data set two

Feature subsets producing the highest predictive values for each individual sensing domain are presented in tables (4.19) - (4.23). Figures (4.17) and (4.18) show a similar trend for features combination in individual sensing domain, as in section (4.4.5), where once the highest merit values of features combination have been attained, any further addition does not necessarily increase the merit score. For CO_2 measurements, AVR_CO_2, FDIFF_CO_2, AF_DIFF_CO_2, AS_DIFF_CO_2 produced the highest merit value of 0.333 for occupancy estimation. Same pattern were obtained for case temperature, PIR, and sound features with highest merit values reaching 0.347, 0.261, and 0.428 respectively. VOC features had the lowest merit score at 0.065.

In evaluating the effectiveness of the individual sensing domain, sound sensors provided the most relevant information for occupancy number estimation followed by case temperature and then CO_2 sensors, figure (4.19). VOC levels displayed insufficient ability for estimation of occupancy numbers estimation.

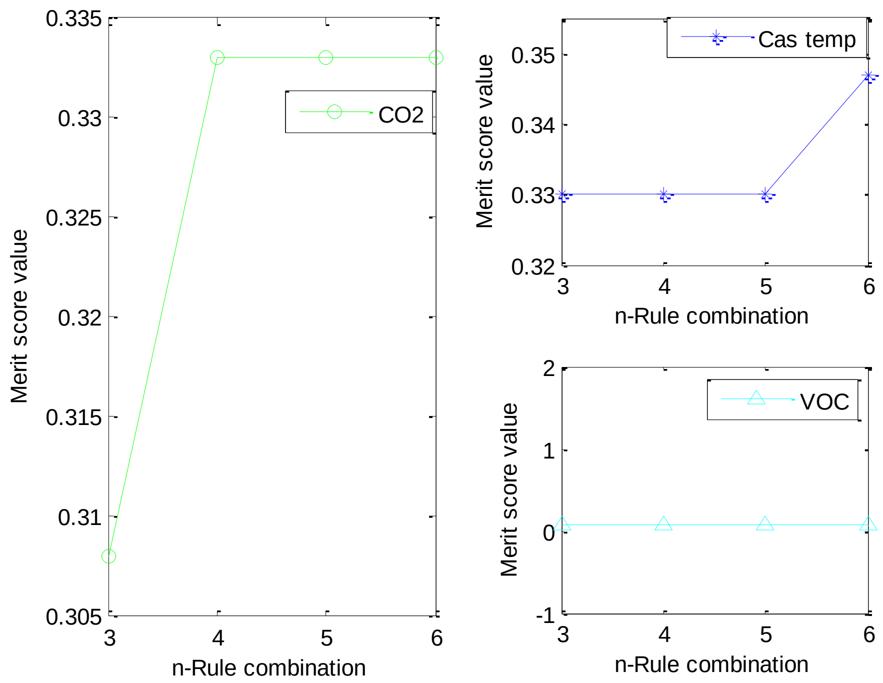


Figure (4.17): n-Rule features combination for individual sensing domain- data set two

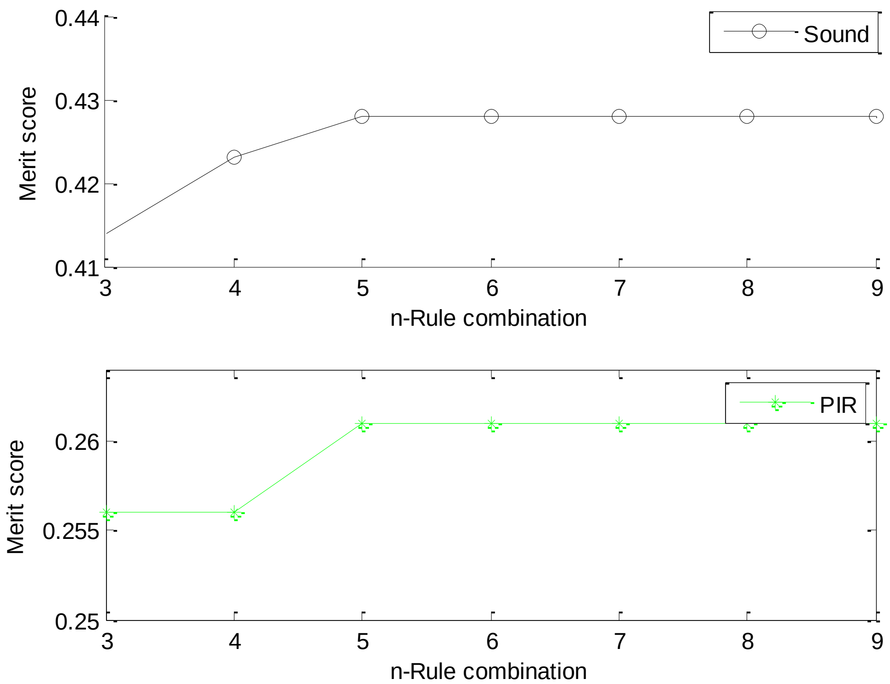


Figure (4.18): n-Rule features combination for PIR and sound sensor - data set two

Table (4.18): Heterogeneous multi-sensory features selection (Scenario one - data set two) indicating the dominant features in colours

n-Rule combination					**Features selected**						**Merit score**
n=1	**FDIFF_CO_2**	**AVR_CAS**	**VHI_LO_PIR**	**VOS_SND**							**0.461**
n=2	**FDIFF_CO_2**	**AVR_CAS**	**VHI_LO_PIR**	**VOS_SND**	**AF_DIFF_CO_2**	**VAR_CAS**	**AF_DIFF_VOC**	**TOS_SND**			**0.478**
n=3	**FDIFF_CO_2**	**AVR_CAS**	**VHI_LO_PIR**	**VOS_SND**	**AF_DIFF_CO_2**	**VAS_CAS**	**AF_DIFF_VOC**	**TOS_SND**	**AF_DIFF_CAS**	**VTONP_SND**	**0.491**
n=4	**FDIFF_CO_2**	**AVR_CAS**	**VOS_SND**	**AF_DIFF_CO_2**	**TOS_SND**	**VTONP_SND**	**VHI_LO_SND**				**0.492**
n=5	**FDIFF_CO_2**	**AVR_CAS**	**VOS_SND**	**AF_DIFF_CO_2**	**TOS_SND**	**VTONP_SND**	**VHI_LO_SND**				**0.492**
n=6	**FDIFF_CO_2**	**AVR_CAS**	**VOS_SND**	**AF_DIFF_CO_2**	**VAS_CAS**	**TOS_SND**	**VHI_LO_SND**	**TLO_HI_SND**	**VLO_HI_SND**		**0.492**

Table (4.19) CO_2 features combination – data set two

n-Rule combination		**Features selected**			**Merit score**
n=3	AVR_CO_2	FDIFF_CO_2			0.308
n=4	AVR_CO_2	FDIFF_CO_2	AF_DIFF_CO_2		0.333
n=5	AVR_CO_2	FDIFF_CO_2	AF_DIFF_CO_2	AS_DIFF_CO_2	0.333
n=6	AVR_CO_2	FDIFF_CO_2	AF_DIFF_CO_2	AS_DIFF_CO_2	0.333

Table (4.20) Case temperature features combination – data set two

n-Rule combination	Features selected			Merit score
n=3	AVR_CAS	FDIFF_CAS		0.330
n=4	AVR_CAS	AF_DIFF_CAS		0.330
n=5	AVR_CAS	AF_DIFF_CAS		0.330
n=6	AVR_CAS	AF_DIFF_CAS	VAR_CAS	0.347

Table (4.21) VOC features combination - data set two

n-Rule combination	Features selected	Merit score
n=3	AVR_VOC	0.065
n=4	AVR_VOC	0.065
n=5	AVR_VOC	0.065

Table (4.22) PIR features combination – data set two

n-Rule combination	Features selected					Merit score
n=3	VHI_LO_PIR	THI_LO_PIR	VOS_PIR			0.256
n=4	VHI_LO_PIR	THI_LO_PIR	VOS_PIR			0.256
n=5	VHI_LO_PIR	THI_LO_PIR	VOS_PIR	TOS_PIR	FDIFF_HI_LO_PIR	0.261
n=6	VHI_LO_PIR	THI_LO_PIR	VOS_PIR	TOS_PIR	FDIFF_HI_LO_PIR	0.261
n=7	VHI_LO_PIR	THI_LO_PIR	VOS_PIR	TOS_PIR	FDIFF_HI_LO_PIR	0.261
n=8	VHI_LO_PIR	THI_LO_PIR	VOS_PIR	TOS_PIR	FDIFF_HI_LO_PIR	0.261
n=9	VHI_LO_PIR	THI_LO_PIR	VOS_PIR	TOS_PIR	FDIFF_HI_LO_PIR	0.261

Table (4.23) Sound features combination – data set two

n-Rule combination	Features selected					Merit score
n=3	TOS_SND	VOS_SND	VTONP_SND			0.414
n=4	TOS_SND	VOS_SND	VTONP_SND	VHI_LO_SND		0.423
n=5	TOS_SND	VOS_SND	VTONP_SND	VHI_LO_SND	TLO_HI_SND	0.428
n=6	TOS_SND	VOS_SND	VTONP_SND	VHI_LO_SND	TLO_HI_SND	0.428
n=7	TOS_SND	VOS_SND	VTONP_SND	VHI_LO_SND	TLO_HI_SND	0.428
n=8	TOS_SND	VOS_SND	VTONP_SND	VHI_LO_SND	TLO_HI_SND	0.428
n=9	TOS_SND	VOS_SND	VTONP_SND	VHI_LO_SND	TLO_HI_SND	0.428

Table (4.24): Features subsets with highest merit score for individual sensing domain- data set two

Individual sensing domain	Features selected						Merit score
CO_2	AVR_CO_2	FDIFF_CO_2	AF_DIFF_CO_2	AS_DIFF_CO_2			0.333
Case temp	AVR_CAS	AF_DIFF_CAS	VAR_CAS				0.347
VOC	AVR_VOC						0.065
PIR	TOS_PIR	VOS_PIR	THI_LO_PIR	FDIFF_THI_LO_PIR	VHI_LO_PIR	AF_DIFF_TLO_HI_PIR	0.262
Sound	TOS_SND	VOS_SND	VTONP_SND	VHI_LO_SND	TLO_HI_SND		0.428

Table (4.25): Heterogeneous multi-sensory features selection (Scenario two - data set two) indicating the dominant features in colours

n-Rule combination	Features selected										Merit score
n=1	FDIFF_CO_2	AVR_CAS	VHI_LO_PIR	VOS_SND							0.461
n=2	FDIFF_CO_2	AVR_CAS	VHI_LO_PIR	VOS_SND	AF_DIFF_CO_2	VAR_CAS	TOS_SND				0.478
n=3	FDIFF_CO_2	AVR_CAS	VOS_SND	AF_DIFF_CO_2	VAR_CAS	TOS_SND	AF_DIFF_CAS	VTONP_SND			0.491
n=4	FDIFF_CO_2	AVR_CAS	VOS_SND	AF_DIFF_CO_2	VAR_CAS	TOS_SND	AF_DIFF_CAS	VTONP_SND	VHI_LO_SND		0.495
n=5	FDIFF_CO_2	AVR_CAS	VOS_SND	AF_DIFF_CO_2	VAR_CAS	TOS_SND	AF_DIFF_CAS	VTONP_SND	VHI_LO_SND	TLO_HI_SND	0.494
n=6	FDIFF_CO_2	AVR_CAS	VOS_SND	AF_DIFF_CO_2	VAR_CAS	TOS_SND	AF_DIFF_CAS	VTONP_SND	VHI_LO_SND	TLO_HI_SND	0.494

PIR sensors may not have produced robust occupancy numbers information, due to placement of one of the sensors, which was mostly triggered by a single occupant, and does not reflect the distribution of occupancy numbers in the space. VOC levels in the space had a poor correlation with occupancy numbers, suggesting occupants may not be the source of VOC levels, and there may be other sources.

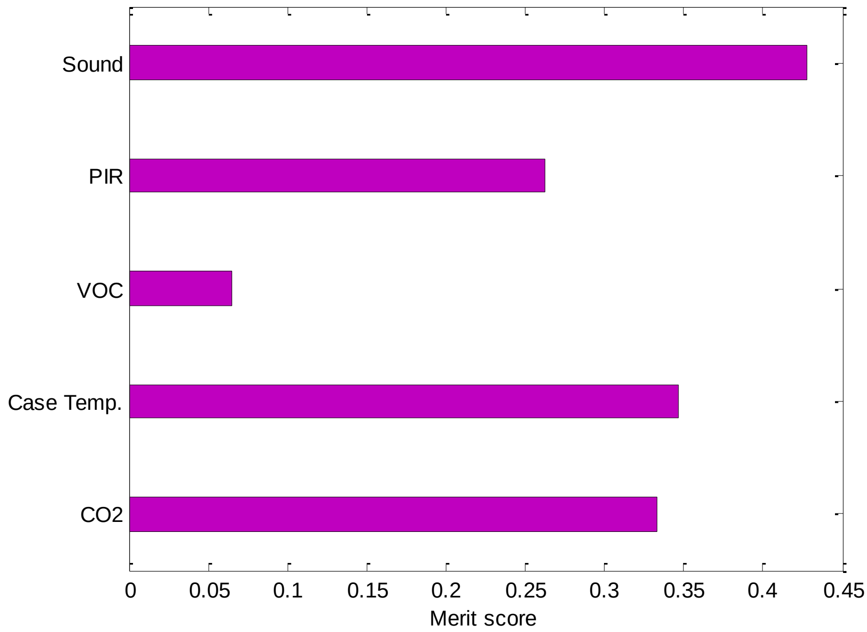


Figure (4.19): Effectiveness of individual sensing domain- data set two

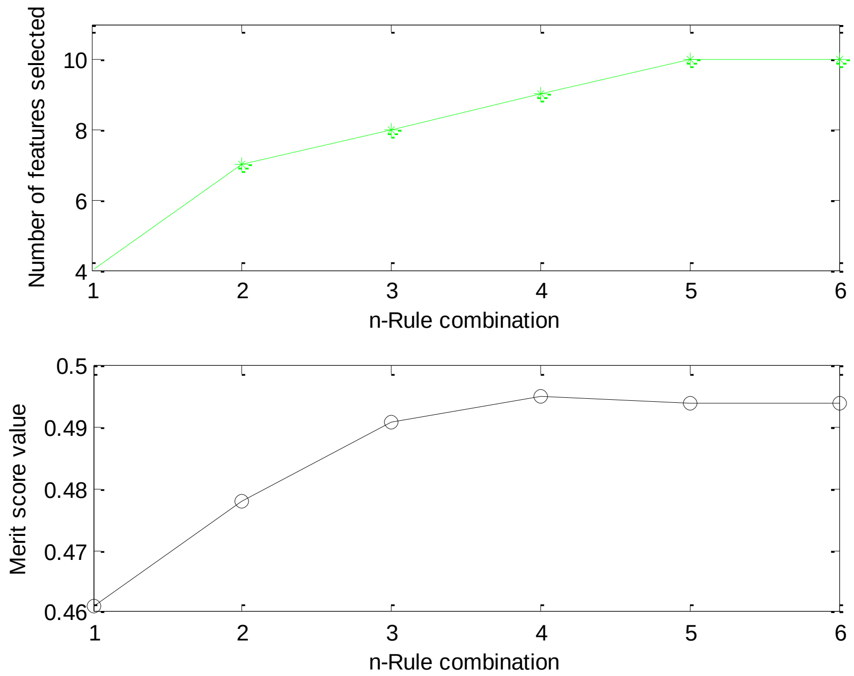


Figure (4.20) Merit score and different n-Rule feature combination

(Scenario two –data set two)

As with the case in scenario one, from figure (4.20) the number of feature combinations increased with merit score, till when n= 5 where the number of selected features was constant, although with an increase in the merit score, suggesting over fitting of data. However, both remain constant at n=6, suggesting the best merit score for the features subsets may have been attained already. From table (4.25), FDIFF_CO_2, AVR_CAS, VOS_SND, AF_DIFF_CO_2, VAR_CAS, and TOS_SND satisfied the adopted 80% selection criterion in the combination process. Following the same process as for data set one, this feature subset was found to have a higher predictive power than the one selected in scenario one, hence was used as inputs for the fusion model.

4.7 Chapter summary

This chapter presented an advanced data processing strategy, capable of facilitating reliable occupancy monitoring in open-plan offices. Various stages in the data processing strategy have been presented, which included pre-processing, features extraction, SU based feature ranking, and correlation-based feature selection.

Features from different sensor types deployed were extracted, and then a feature ranking analysis was carried out to establish the correlation of features with occupancy levels. Once ranks of the features had been determined, candidate features were selected using correlation based feature selection filter implemented in *WEKA.* A key finding from the features selection process is that sound features were consistently dominant in the data sets analysed. Features selected tend to be dependent on the environmental dynamics in the observed environment. This was quite evident in data set one, where sound and motion features were dominant.

CHAPTER 5

SENSOR FUSION FOR OCCUPANCY ESTIMATION

5.0 Introduction

In this chapter, candidate sensor features established in chapter four are combined using a machine learning model for reliable estimation of occupancy levels. Experimental results for building occupancy monitoring using various sensor network configurations are presented. Tests were conducted to demonstrate, validate and assess the effectiveness of the proposed occupancy estimation approach. An assessment was carried out in an open-plan office space using low-cost and non-invasive sensors. This chapter is organised as follows; Section 5.1 conceptualizes the occupancy estimation problem. Section 5.2 introduces metrics used in evaluating the model performance. Section 5.3 describes the structure of the neural network used in the analysis. Section 5.4 presents occupancy estimation results using redundant and heterogeneous multisensory networks, while 5.5 presents a comparison of various sensors network configurations. Section 5.6 examines the use of all sensory features (including redundant and irrelevant features) against selected multisensory features. Section 5.7 studies the concept of resilience in an occupancy sensing network and section 5.8 looks into the feasibility of using alternative sensors for ventilation control. Section 5.9 compares model estimations using different machine learning approach whilst section 5.10 examines the generalisation ability of a learned model for occupancy level estimations in different spaces. Lastly, section 5.11 presents results for weekend estimations while 5.12 presents the chapter summary.

5.1 Problem formulation

Consider an open-plan office with numerous monitoring points, such as in figure (3.4) and (3.7) in section (3.1.1). To better control HVAC and electrical systems, occupancy levels in the space is of interest. Suppose there are $\boldsymbol{b}$ sensors deployed in an open-plan office which measure indoor environmental variables for the purpose of estimating the number of occupants. Let $\boldsymbol{m_b}$ be the actual occupancy number in the test area observed a system, and $\widehat{\boldsymbol{m_b}}$ be the number of occupants estimated from sensor measurements**.** Note that it is possible that $\boldsymbol{m_b} \neq \widehat{\boldsymbol{m_b}}$. For instance, in this research, the infrared camera used for ground truth occupancy count may report 10 occupants in the space, i.e., $\boldsymbol{m_b}$ =10. If on the average only 70% of occupants are detected by the sensor network deployed, then $\widehat{\boldsymbol{m_b}}$ = 7. Define $\boldsymbol{m_o}$= $(\boldsymbol{m_1} \dots . \boldsymbol{m_b})^{\boldsymbol{\tau}}$ and $\boldsymbol{m_e}$ = $(\widehat{\boldsymbol{m_1}} \dots . \widehat{\boldsymbol{m_b}})^{\boldsymbol{\tau}}$, where $\boldsymbol{m_o}$ denote the vector of observations by the infrared camera, and $\boldsymbol{m_e}$ is the vector of estimations of $\boldsymbol{b}$ sensors, and τ is the transpose. Given $\boldsymbol{m_o}$ **and** $\boldsymbol{m_e}$**,** the task is therefore to find an estimate $\widehat{\boldsymbol{m_b}}$ to achieve a minimum error, i.e.,

$$\frac{\boldsymbol{min}}{\boldsymbol{m_e} \in \overline{\boldsymbol{Z^-}}} \boldsymbol{E}[(\boldsymbol{m_e} - \boldsymbol{m_o})^2 | \boldsymbol{m_o} - \boldsymbol{i}] \tag{5.1}$$

Where $\overline{\boldsymbol{Z^-}}$ =$\boldsymbol{Z^+} \cup \{\boldsymbol{0}\}$ is a set of non-negative integers, $\boldsymbol{i}$= $(\boldsymbol{i_1}, \dots . \boldsymbol{i_b})^{\boldsymbol{\tau}}$is a constant vector, *E* is the error vector and *min* is the minimum.

Hence, with candidate features identified, the next step is implementing a proper fusion strategy for occupancy estimation. In general, multi-sensor fusion strategies depend significantly on correlations between sensor features and the predictive class, which has been clearly elucidated in chapter four (section 4.5). The Dasarathy fusion model (Dasarathy, 1997) has been adopted as the fusion strategy, see figure (5.1). The objective of this strategy is to combine information from candidate features from a variety of sensors to derive occupancy information that may not be obtainable from a single sensor type. Before presenting the occupancy estimation model applied in this thesis, the metrics applied to evaluate the model performance are presented in the next section.

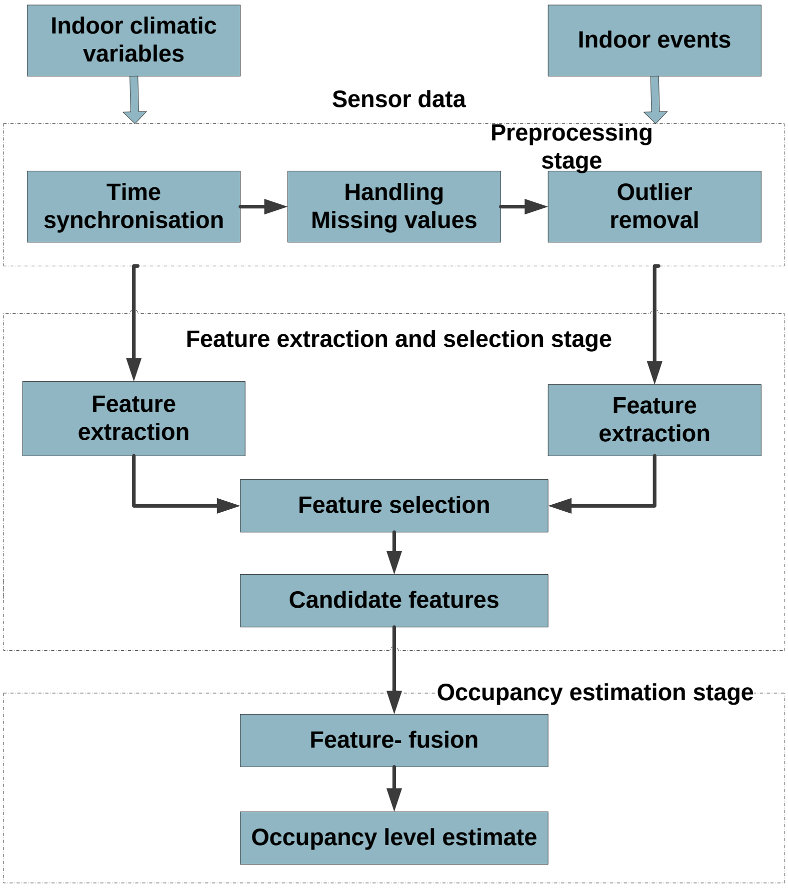


Figure (5.1): Basic approach of the fusion strategy implemented in the thesis

5.2 Model performance metrics

The test results are evaluated to provide occupancy information, such that HVAC systems can be proactively adjusted based on it. In this thesis, occupancy detection does not indicate the rate of sensing occupancy presence (e.g. occupied and unoccupied), instead it refers to the rate of sensing occupancy numbers in the test area. In the occupancy detection literature, Root Mean Square Error (RMSE) is mostly used to evaluate the performance of a learned occupancy detection model. However, in order to carry out a robust error analysis, various standard statistical performance evaluation measures, i.e. RMSE, the Mean Absolute Percentage Error (MAPE), coefficient of determination (R^2), and relative absolute error (RAE) were employed to validate each model's performance. These are commonly applied, and are representative techniques for model performance evaluation in machine learning research.

- **Root mean square error (RMSE)**

RMSE measures the difference between estimated occupancy and actual occupancy data. This gives an indication of how close the model estimations are fitted or close to the actual occupancy data. This metric seeks to aggregate into a single measure of predictive power, the sum of the squared errors of the entire model estimations. Lower RMSE values indicate better model performance. RMSE was computed using equation (5.2).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Q_E(i) - Q_O(i))^2} \tag{5.2}$$

Where Q_E = Model estimations

Q_O= Actual occupancy data

n= Total data instance

- **Mean absolute percentage error (MAPE)**

MAPE which gives the model accuracy was used to make a term-by-term comparison of the relative error in the model estimations with respect to actual occupancy data. MAPE is an unbiased metric for measuring the predictive capacity of a model. It gives an indication of the model accuracy in a fitted time series value in statistics, and it is expressed in generic percentage terms. Wang et al. (2009) have applied MAPE to assess the performance of a hydrological model using three machine learning techniques. It makes more sense to apply MAPE only to measurements during the occupied periods than otherwise. Its formula is given as in equation (5.3), while the model accuracy was computed as in equation (5.4). The model accuracy has been reported to two decimal places, as a PID controller in a building would normally require this level of precision for robust systems control.

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{Q_E(i) - Q_O(i)}{Q_O(i)}\right| \mathrm{X}\ 100 \tag{5.3}$$

$$\text{Model accuracy} = 100\% - \mathrm{MAPE} \tag{5.4}$$

- **Relative absolute error (RAE)**

RAE measures the total absolute error and normalizes it by dividing with the total absolute error of a simple estimation model (Armstrong and Collopy, 1992). It gives an indication of the relative error of a simple estimation model. For a perfect fit, the numerator is equal to 0 and RAE = 0. RAE ranges from 0 to infinity, with smaller values indicating better model performance. RAE was computed as given in equation (5.5)

$$RAE = \frac{\sum_{i=1}^{n} \left| Q_{E(i)} - Q_{O(i)} \right|}{\sum_{i=1}^{n} \left| Q_{O(i)} - \bar{Q}_O \right|} \tag{5.5}$$

$$\bar{Q}_O = \frac{1}{n} \sum_{i=1}^{n} Q_{O(i)} \tag{5.6}$$

Where $\bar{Q}_O$ is the mean of the actual occupancy data.

- **Coefficient of determination (R^2)**

Coefficient of determination measures the goodness of fit of a model. This metric gives an indication of the proportion of variance (fluctuation) between model estimations and actual occupancy data. It measures how well a regression line represents the data. It is evaluated as a number between 0 and 1.0. As the R^2 value tends to 1.0, the closer the model estimations is to actual occupancy data. R^2 was computed using equation (5.7). Shiri and Kiri (2011) and Bilgehan (2011) have also applied R^2 for performance evaluation of various machine learning models.

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \tag{5.7}$$

$$SS_{res} = \sum_{i=1}^{n} \left(Q_{O(i)} - Q_{E(i)} \right)^2 \tag{5.8}$$

$$SS_{tot} = \sum_{i=1}^{n} \left(Q_{O(i)} - \bar{Q}_{O(i)} \right)^2 \tag{5.9}$$

Where SS_{res} is the total sum of squared errors, SS_{tot} is the total sum of squares

RMSE measure is the most sensitive of all the measures examined, especially for outliers, as it gives a disproportionate weight to such errors because it is a squared quantity, whereas RAE is less sensitive to large error (Armstrong and Collopy, 1992). MAPE is biased to data with low model estimation values tending to zero. For R^2, where SS_{tot} is larger than SS_{res}, the model performance becomes unclear.

5.3 Neural network based fusion for occupancy levels estimation

A back propagation artificial neural network (NN) was applied for occupancy level estimation analysis. The model was implemented using the *MATLAB* Neural Network toolbox. Figure (5.2) shows the NN architecture used for the study. Generally, inputs for the NN model were candidate features as informed from the feature selection process provided in chapter four. There is lack of any established theory for specification of how many hidden layers or the number of neurons that is required for approximation of a given function. Various NN architectures were tried by the researcher, and an appropriate model structure was adopted when the error variation between estimated and actual occupancy data became sufficiently small. The Log Sigmoid transfer function was used in both hidden layers, while a linear function was used in the output layer. 15 neurons were used in each of the connecting layers, with other parameters such as a learning rate of 0.05, number of epochs of 500 and momentum of 0.9. The training phase is repeated for 10 times to increase the probability of reaching a global solution. The resulting average from the outputs of the training phases was used for analysis in the next stages.

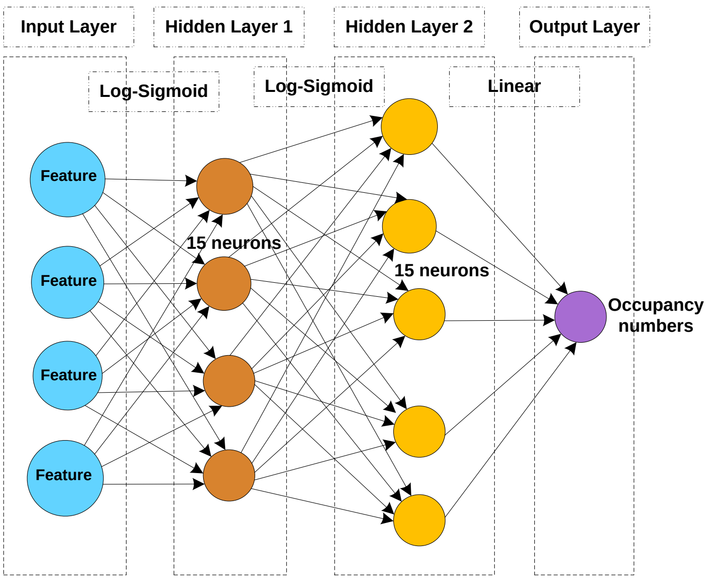


Figure (5.2): Architecture of the neural network applied for fusion

5.4 Individual sensing network and heterogeneous multisensory network

In order to investigate the use of individual sensing network, features extracted from sensors as shown in table (4.24) were used as model inputs and tested for occupancy level estimation. Results are presented for PIR, case temperature, CO_2, and sound sensor network in preceding sections (5.4.1- 5.4.4). Occupancy profile in the area under test often showed variability between different days. Using data set two from test-area three, which covered a continuous time period of November 27- December 20, 2012, results for a typical week (excluding weekend days) are presented and compared against each sensing network. In this analysis, data from the first two weeks were used for model training, while data from the remaining week were used for model testing.

5.4.1 PIR sensor network

Studies have shown that a PIR sensor network can work well for occupancy presence detection for building controls (especially lighting control), as opposed to single point PIR occupancy sensing, as it tends to minimise the incidence of false-offs, and therefore improve detection accuracy (Tiller et al., 2010). Current PIR systems cannot differentiate between one or more occupants in an observed space, although a precise count would be useful for energy management. This analysis examines the use of a PIR sensor network for occupancy numbers detection. Six features obtained from all three PIR sensors deployed in the test area (TOS_PIR, VOS_PIR, THI_LO_PIR, FDIFF_THI_LO_PIR, VHI_LO_PIR, AF_DIFF_TLO_HI_PIR), shown in table (4.24) were used as inputs for the NN based fusion model. Figure (5.3) and (5.4) show estimated and actual occupancy data for two typical days in a week. Occupancy levels were estimated with an RMSE of 1.683 on the 18/12/2012, and RMSE of 1.432 on the 19/12/2012, with low accuracies of 43.23% and 50.17% during occupied period respectively. However, such low accuracy for during occupied period is not surprising; as a single occupant can trigger all sensors in the space while walking across the room, suggesting that there may be more occupants (since some of the features captures the sensor's pulse rates). Again, these sensors fail to detect minor motion and can be sensitive to temperature changes in the observed space. The model produced good results during unoccupied periods. Model estimations showed significant variation from the actual occupancy data as indicated by the poor R^2 value of 0.503 and relative error of 0.491 for 18/12/2012, while an R^2 value of 0.634 and relative error of 0.390 for 19/12/2012 was recorded.

An average RMSE of 1.518, and accuracy of 50.50% for occupied periods was achieved using this sensory network, see table (5.1). The average relative error and R^2 was 0.457 and 0.568 respectively. These suggest poor predictive ability of a PIR sensor network for occupancy numbers detection, especially during occupied periods, model estimations tend to be rather noisy. It is clear that this network may be more appropriate for presence detection than otherwise. Further work may be necessary to use this network to distinguish the number of occupants with improved accuracy.

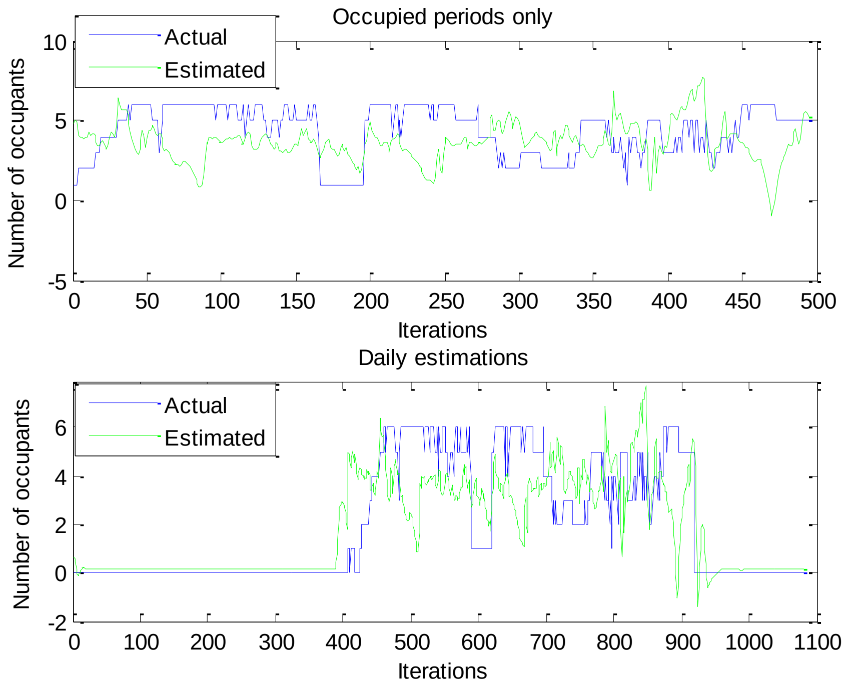


Figure (5.3): Occupancy results using PIR sensor network on 18/12/2012

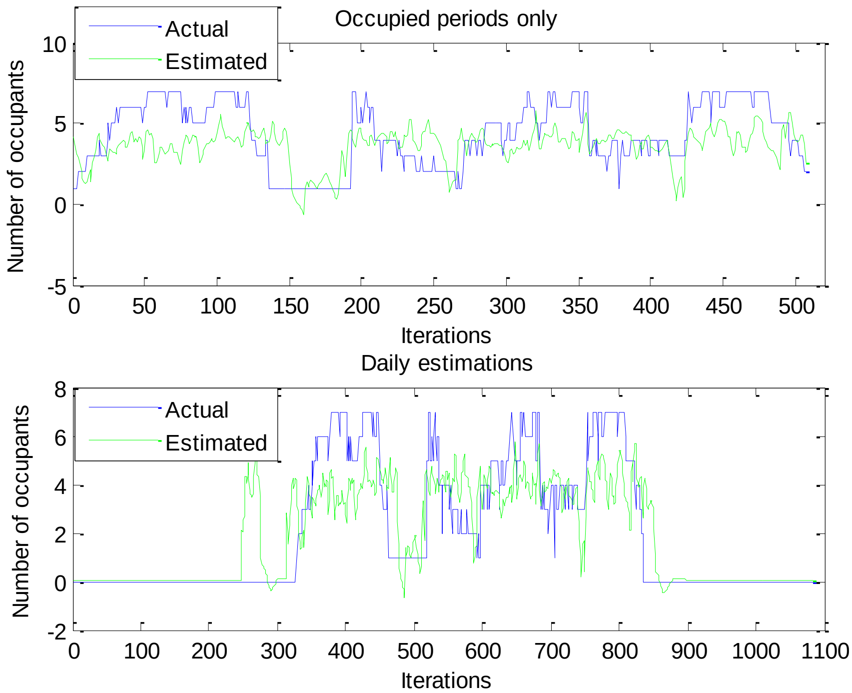


Figure (5.4): Occupancy results using PIR sensor network on 19/12/2012

Table (5.1): Model performance for a typical week using a PIR sensor network

Week days	RMSE	R^2	RAE	Accuracy
14th	1.313	0.569	0.484	49.71
17th	1.945	0.402	0.579	41.08
18th	1.683	0.503	0.491	43.23
19th	1.432	0.634	0.390	50.17
20th	1.216	0.732	0.341	68.33

5.4.2 Case temperature sensor network

In most modern office buildings, each occupant work station is often fitted with a computer. The use of a case temperature monitoring network may be useful for occupancy level estimation in open-plan office spaces. Although, these are stand-alone loggers, the principle is the same for a full RF sensor network. For DCV control strategies based on occupancy patterns, usage patterns of electrical appliances can provide insights into physical space use, forming control strategies for ventilation. Features such as AVR_CAS, AF_DIFF_CAS, and VAR_CAS, shown in table (4.24), which produced the highest predictive capacity were used as the model inputs for occupancy level estimation.

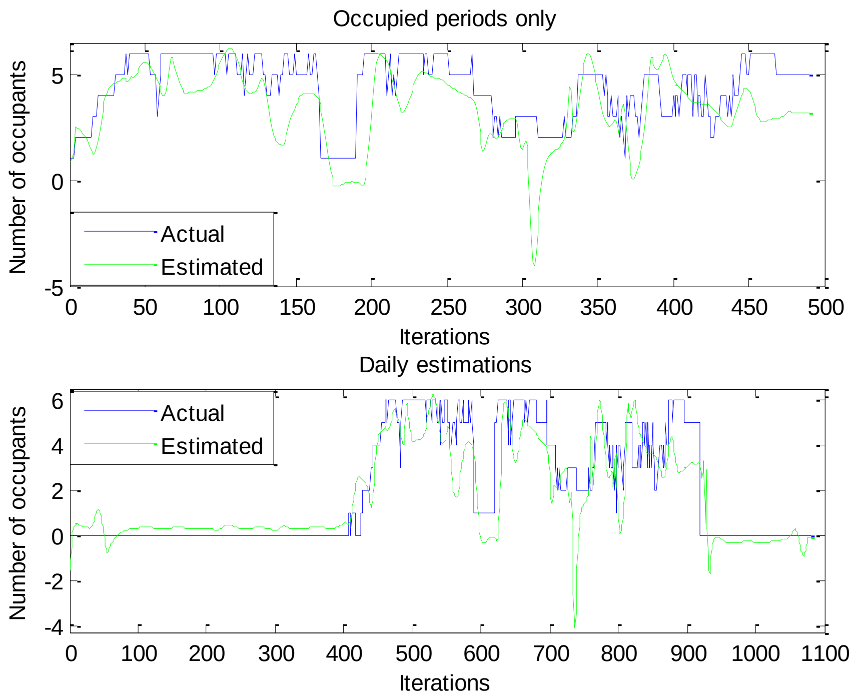


Figure (5.5): Occupancy results using case temperature sensor network on 18/12/2012

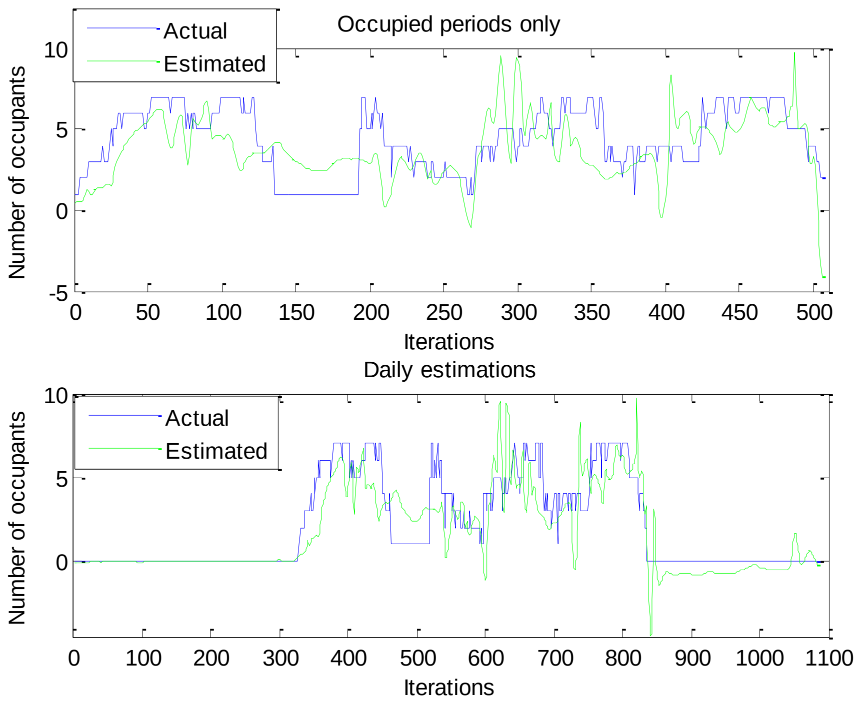


Figure (5.6): Occupancy results using case temperature sensor network on 19/12/2012

From figure (5.5) and (5.6), model estimations track fairly well with actual occupancy data, with an RMSE of 1.141 and an accuracy of 62.58%, during occupied instances on the 18/12/2012. On the 19/12/2012, actual occupancy level increased by 1 from the typical 6 occupants accommodated in the office. Model estimations show a larger variance with actual occupancy data than the previous day with an R^2 value of 0.570, RMSE of 1.346 and accuracy of 54.71% on the 19/12/2012. This low accuracy may be in part that while this network may provide a good indication of occupancy numbers, when all workstation computers in the observed space are instrumented for case temperature monitoring, it is unable to detect people not using a computer. As presented in table (5.2), average RMSE of 1.334, R^2 value of 0.598, RAE of 0.407, and accuracy 55.70% was achieved for the week examined. Although, the network shows relatively poor performance for occupancy numbers tracking, it can be useful for establishing appliance usage (Brown et al., 2011). Such information can be useful for developing an occupancy-driven power management strategy for office computers, thus providing opportunities for energy saving. In cases, where occupants forget to turn off computers during or after work hours, this sort of network may be a useful intervention to indicate space vacancy, thereby facilitating powering down of computers.

Table (5.2): Model performance for a typical week using a case temperature sensor network

Week days	RMSE	R^2	RAE	Accuracy
14th	1.601	0.427	0.552	44.67
17th	1.352	0.581	0.414	56.57
18th	1.141	0.725	0.331	62.58
19th	1.346	0.570	0.378	54.71
20th	1.231	0.688	0.360	60.00

5.4.3 Sound sensor network

An in-depth literature review strongly suggests that a sound sensing network has not been applied for occupancy numbers estimation. The effectiveness of this network for occupancy number estimation is examined by combining features such as

TOS_SND, VOS_SND, VTONP_SND, VHI_LO_SND, TLO_HI_SND, shown in table (4.24). Table (5.3) presents the performance metrics for the sound sensor network for a typical week. An average RMSE of 1.225, R^2 value of 0.676, RAE of 0.371, and accuracy 60.60% achieved, suggesting the model estimation were fairly close to the actual occupancy data.

Table (5.3): Model performance for a typical week using a sound sensor network

Week days	RMSE	R^2	RAE	Accuracy
14^{th}	1.304	0.585	0.452	54.07
17^{th}	1.205	0.662	0.349	60.61
18^{th}	1.176	0.697	0.352	60.26
19^{th}	1.306	0.656	0.376	58.71
20^{th}	1.132	0.778	0.328	69.34

Figure (5.7) and (5.8), show model estimations and actual occupancy data on the 18/12/2012 and 19/12/2012 respectively. The network was effective for occupancy monitoring with negligible variations, between model estimations and actual occupancy levels, during unoccupied times. However, both data were in good sync for occupied periods. An RMSE of 1.176 and an accuracy of 60.62% were recorded on the 18/12/2012. On the 19/12/2012, an RMSE of 1.306 and an accuracy of 58.71% were recorded. These results are promising, given that the sensors are fairly simple (in operation). Besides, sound measurements can be subject to interference from sources other than occupants which can easily degrade the model performance. Use of these low-cost sound sensors clearly holds potential for occupancy numbers monitoring.

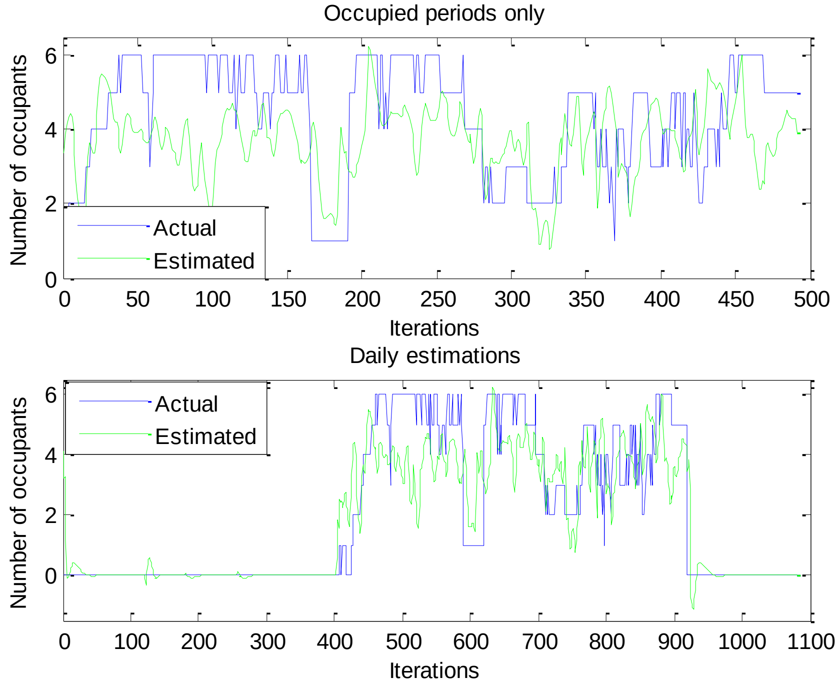


Figure (5.7): Occupancy results using sound sensor network on 18/12/2012

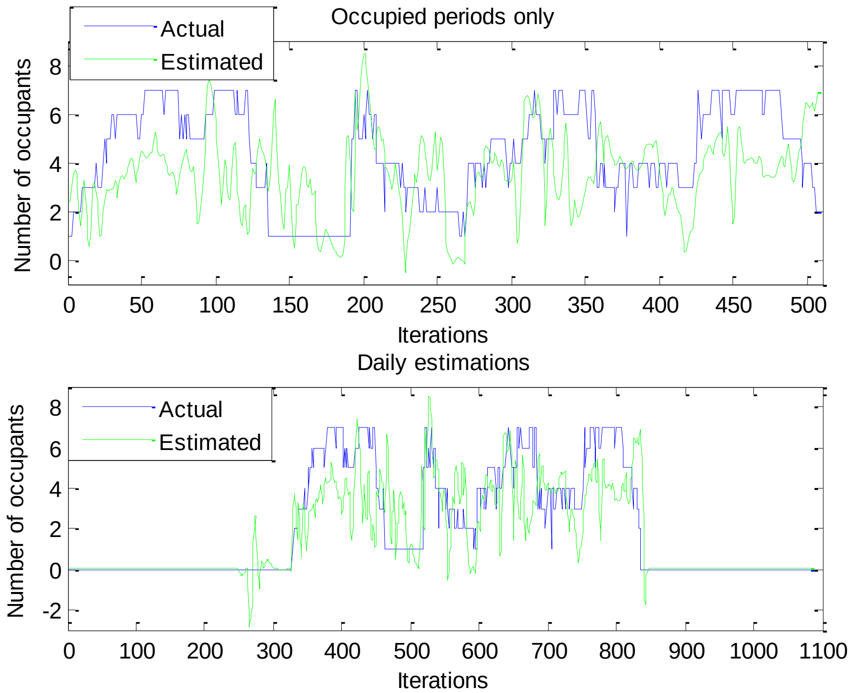


Figure (5.8): Occupancy results using sound sensor network on 19/12/2012

5.4.4 Carbon dioxide (CO_2) sensor network

CO_2 sensing networks are usually deployed for ventilation control operations in many buildings. These sensors monitor CO_2 levels as against a particular threshold as per standards for system control (ASHRAE, 2004), and only provide an abstracted indication of occupancy number count. Such a network may be an expensive option for occupancy level monitoring, with a typical commercial brand (such as the Telaire 7001 series CO_2 sensor) implemented for ventilation control in many office buildings costing as much as £350.00 (GE-Sensing), although BEMS CO_2 sensors may cost around £200.00. Features such as AVR_CO_2, FDIFF_CO_2, AF_DIFF_CO_2 and AS_DIFF_CO_2 in table (4.24), were combined to examine the use of this network for occupancy level estimation. Table (5.4) shows the network performance in a typical week.

Table (5.4): Model performance for a typical week using a CO_2 sensor network

Week days	RMSE	R^2	RAE	Accuracy
14th	0.863	0.837	0.256	71.08
17th	1.255	0.634	0.384	58.54
18th	1.120	0.707	0.366	59.78
19th	1.137	0.783	0.340	61.55
20th	0.880	0.821	0.273	70.92

It was difficult to establish a reliable threshold for CO_2 concentration levels with corresponding occupancy numbers, as CO_2 levels in the space are affected by various factors including air infiltration (caused by window opening) and outdoor CO_2 levels. However, with an average RMSE of 1.067, R^2 value of 0.756, RAE of 0.324, and accuracy 64.37% achieved for the week analysed, model estimations track well with actual occupancy data, suggesting occupants may be largely responsible for CO_2 levels in the space. Although, due to slow CO_2 decay rates in the space, model results show rather noisy estimates during unoccupied times as in figures (5.9) and (5.10).

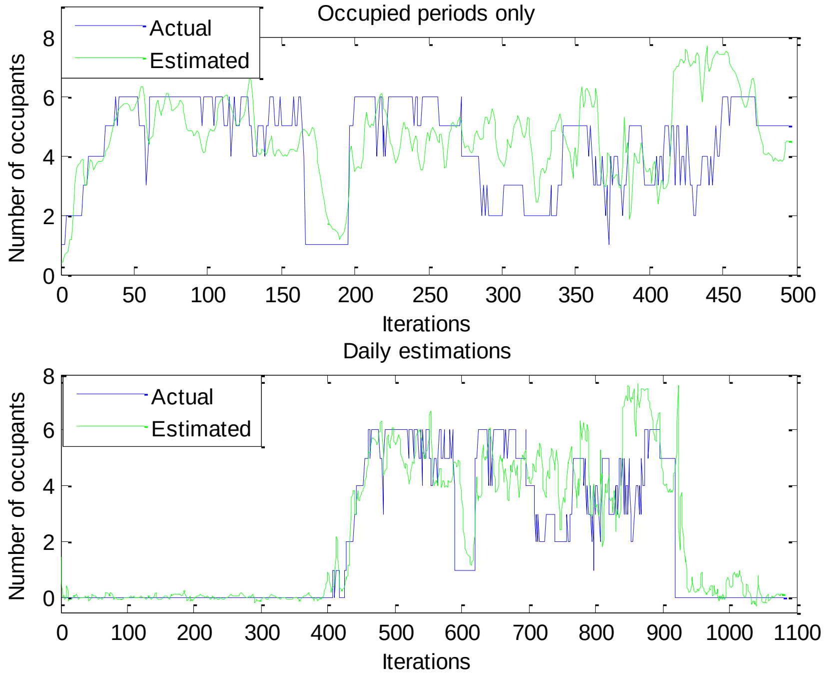


Figure (5.9): Occupancy results using CO_2 sensor network on 18/12/2012

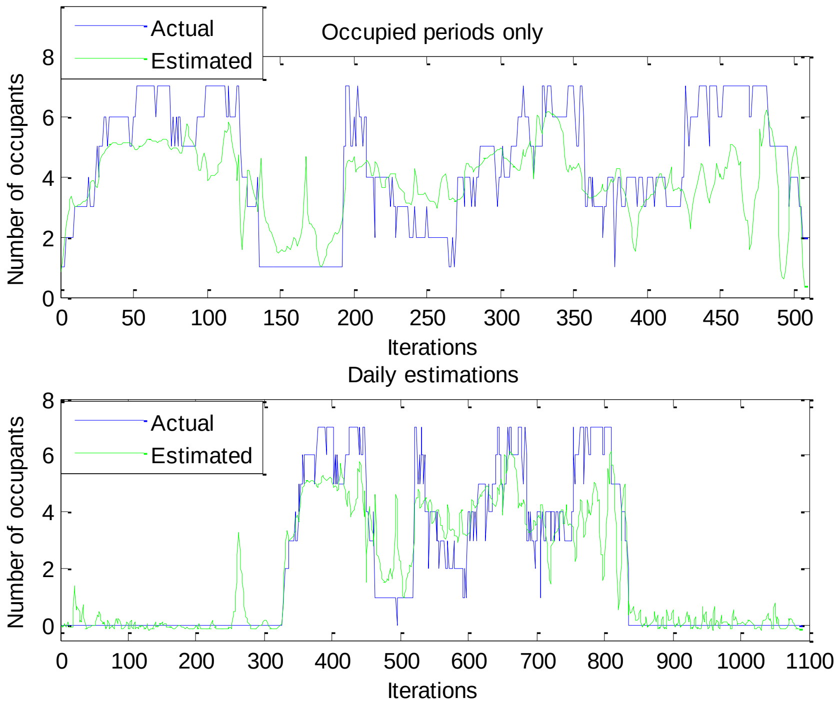


Figure (5.10): Occupancy results using CO_2 sensor network on 19/12/2012

5.4.5 Heterogeneous multi-sensory network

The use of a multi-sensory network comprising of different sensor types for building occupancy sensing is relatively new. The complementary nature of a heterogeneous sensor fusion model can harness the strength of sensors while reducing the impact of their weakness, facilitating better model performance (Hall and Llinas, 2001). However, this is subject to the quality of data used in the fusion process. For this network, features (such as FDIFF_CO_2, AF_DIFF_CO_2, AVR_CAS, VAR_CAS, VOS_SND and TOS_SND) in table 4.25 were combined for occupancy levels estimation.

Figures (5.11) and (5.12) show model estimations and actual occupancy data for two typical days. It is clear from these plots that model estimations are in good sync with actual occupancy numbers, with an RMSE of 0.815 on the 18/12/2012, and RMSE of 1.064 on the 19/11/2012. The model particularly performs well during unoccupied periods; this is not surprising as measured indoor climatic variables do not show any significant temporal variation. Although, for some days in the week examined during unoccupied period, model estimations indicated there were occupants in the space. This may be due to the slow CO_2 decay rates in the test area, which can sometimes take till about 8:00am the next morning for CO_2 levels to completely decay.

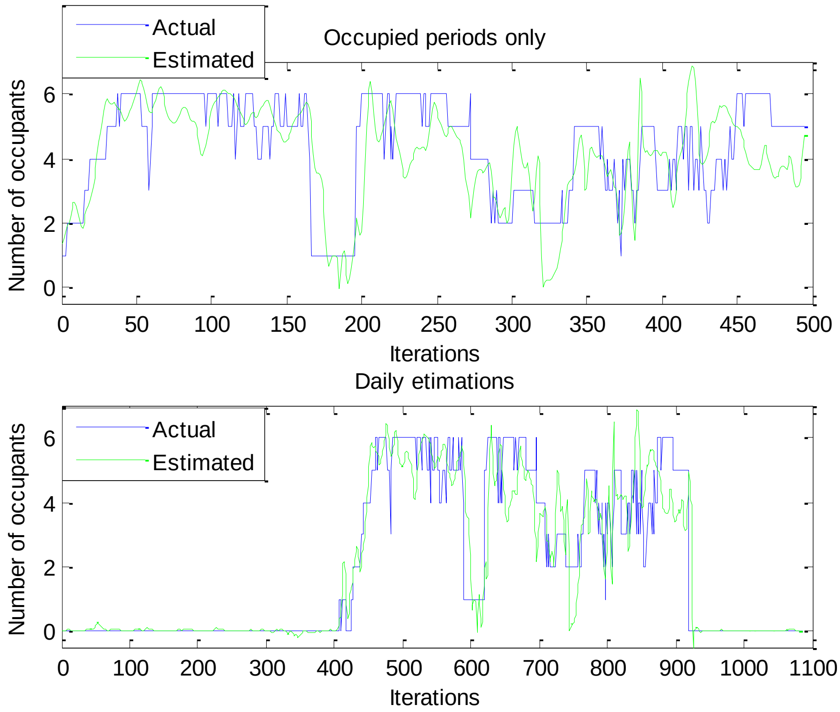


Figure (5.11): Occupancy results using heterogeneous multi-sensor network on 18/12/2012

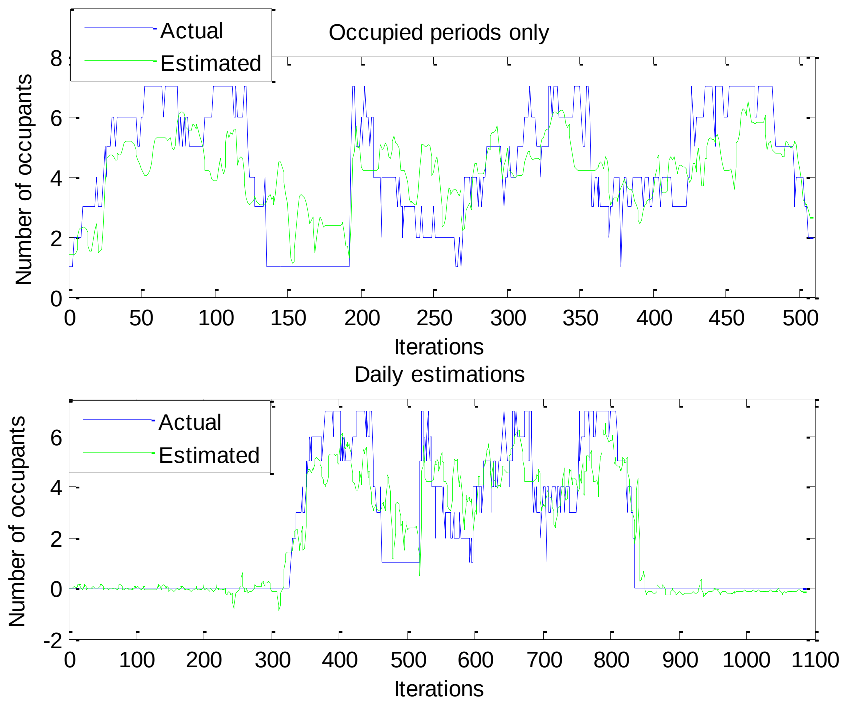


Figure (5.12): Occupancy results using heterogeneous multi-sensor network on 19/12/2012

During occupied periods, model estimation accuracy reached 74.67%, showing close tracking with ground truth data. Although, there are certain days where accuracy was relatively low (e.g. 62.24% on the 17/12/2012). This shows that some amount of occupancy variation can occur between days. Table (5.5) presents the model performance for a typical week. Although, model outputs are in decimal formats and may not represent practical observations i.e. number of occupants cannot be 4.12, hence model outputs may require quantisation. However, the outputs are still useful for occupancy driven HVAC systems, since certain level of error is acceptable. Besides, HVAC systems are not traditionally viewed as needing to be very sensitive, such that they respond to slight changes in occupancy numbers. For instance, a change in the number of occupants by one, normally would not cause any significant HVAC operation, unless the space switches from occupied to unoccupied, and vice versa.

Overall, the model sometimes struggles when there are abrupt changes in occupancy levels, which again may be linked to the slow CO_2 decay rates. In addition, CO_2 sensors are slow to detect incoming occupants. For occupancy driven HVAC control operations, this may not have any significant ramification, as the system is not expected to produce a control action for abrupt occupancy changes, or short occupancy durations.

Table (5.5): Model performance for a typical week using a heterogeneous multi-sensor network

Week days	RMSE	R^2	RAE	Accuracy
14^{th}	0.842	0.849	0.244	73.20
17^{th}	1.161	0.706	0.349	62.24
18^{th}	0.815	0.859	0.229	74.67
19^{th}	1.064	0.827	0.288	68.53
20^{th}	0.845	0.830	0.246	71.89

5.5 Occupancy estimation using various sensing network configurations.

In general, the RMSE measure increased with RAE, and varied inversely with R^2 and the model accuracy for the week examined. In figure (5.13), the best average daily testing RMSE of 0.945 was achieved using a heterogeneous combination of sensors. This is considered good, since the number of occupants varied mostly between 0 and 6. This suggests that the model estimations are usually within 1 of the actual occupancy number. Overall, PIR sensor network showed the largest RMSE values for days in the week, with values reaching 1.945; this was followed by the case temperature network with a peak RMSE of 1.601. Both sound and CO_2 sensor network showed good performance with their RMSE not exceeding 1.255.

Figure (5.14) shows the accuracy of the test results for different days in a typical week using various sensor network configurations between 14/12/2012 and 20/12/2012. Different accuracies for days of the week were recorded, with variation between the highest and lowest values reaching up to 27.25% for PIR sensor network, 17.91% for case temperature sensor network, 15.27% for sound sensor network, 12.54% for CO_2 sensor network and 12.43% for the heterogeneous multi-sensory network. These variations may be influenced by the occupancy dynamics for a particular day. However, a heterogeneous multi- sensory network produced the least variation in the model daily performance.

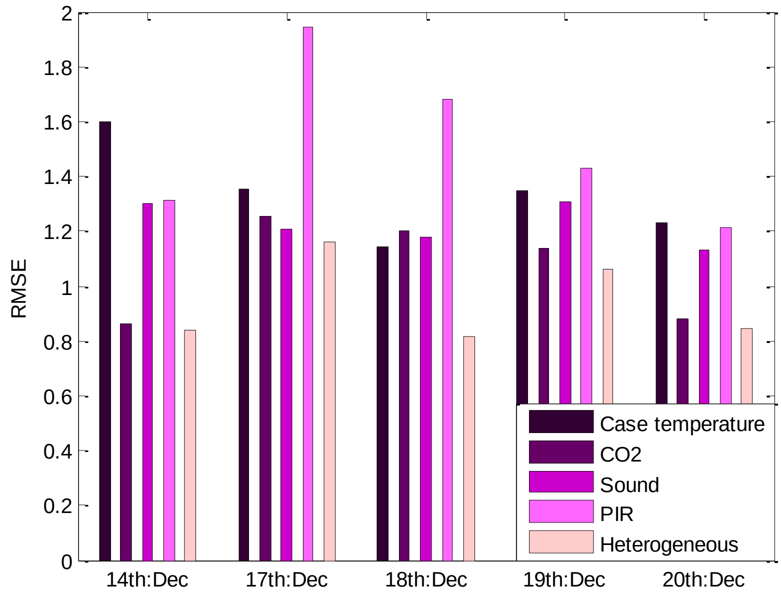


Figure (5.13): RMSE values for various sensing network configurations in a typical week

`

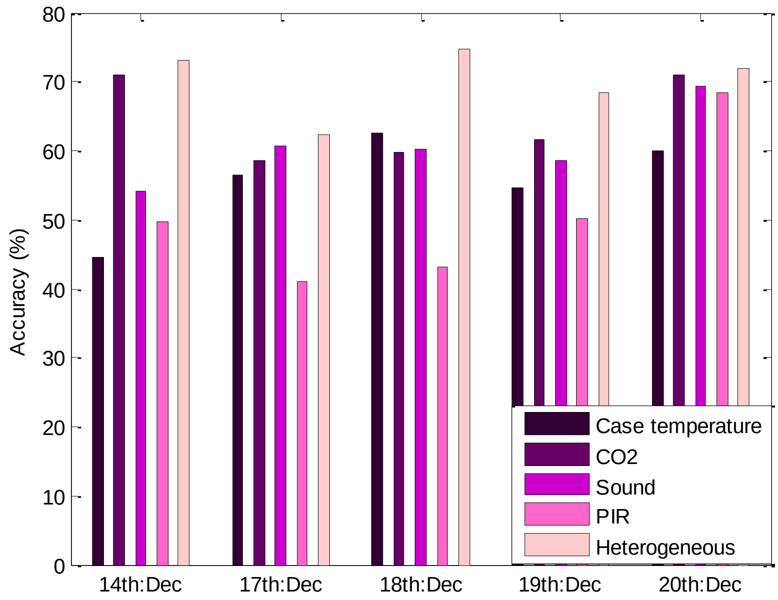


Figure (5.14): Occupied times accuracy for various sensing network configurations

From figure (5.15) and (5.16), the heterogeneous multi-sensory network configuration produced the least relative error and variations between model estimations and actual occupancy data in weekdays examined. On the 14/12/2012, CO_2 and multisensory network show similar relative errors, at 0.256 and 0.244 respectively, and their R^2 values at 0.837 and 0.849 also vary accordingly. Other sensory configurations showed poor RAE and R^2 values relative to both. This trend was similar for 19/12/2012 and 20/12/2012. However, on the 17/12/2012 and 18/12/2012, the CO_2 sensor network showed larger relative error and lower R^2 values compared to other days in the week. This may be due to noisy model estimates recorded during unoccupied times caused by slow CO_2 decay rates in the space. Sound network with an RAE of 0.349 and R^2 of 0.662 performed better than CO_2 sensor network on the 17/12/2012. While case temperature network outperformed sound and CO_2 network on the 18/12/2012, with a RAE of 0.331 and R^2 value of 0.725. Overall, PIR sensor network showed the poorest performance in the week, only performing better than case temperature network on the 14/12/2012 and 20/12/2012.

In summary, a heterogeneous multi-sensory network clearly outperformed others, as indicated by the metrics applied for assessing performance of different network configurations. The CO_2 sensor network also shows good performance, although CO_2 decay rates in a space may impact on the model estimations and control strategies, especially during unoccupied times. The use of this PIR sensor configuration produced the worst performance compared to others. Case temperature monitoring is useful for occupancy monitoring, although further testing is recommended using bigger data samples collected from various office spaces.

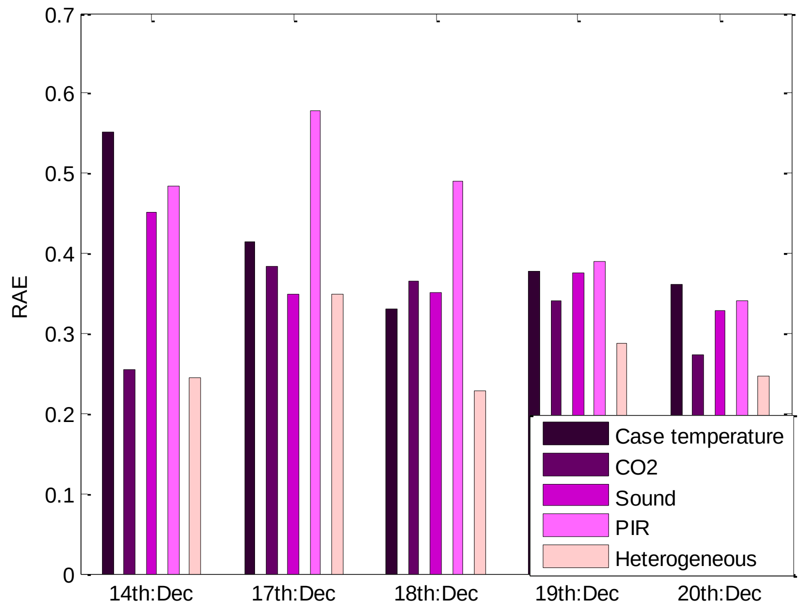


Figure (5.15): RAE values for various sensing network configurations in a typical week

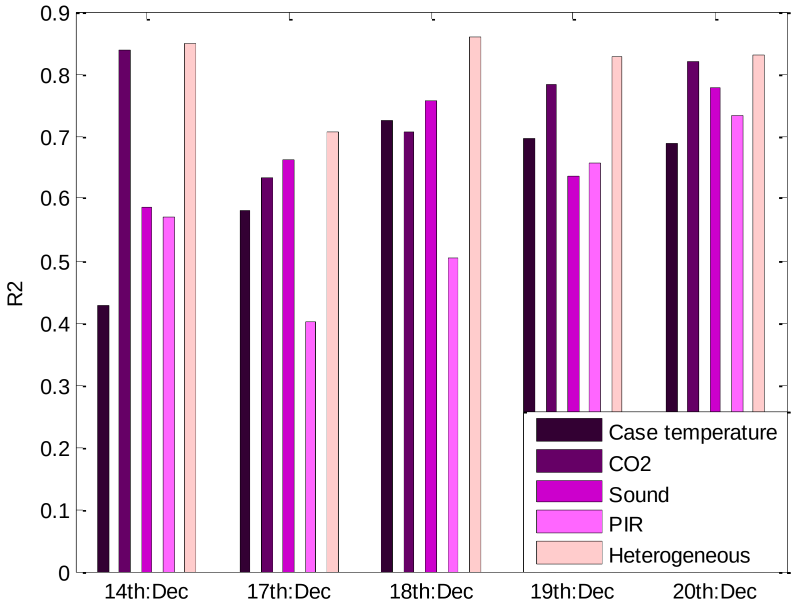


Figure (5.16): R^2 values for various sensing network configurations in a typical week

5.6 Complete set of sensory inputs and optimal multi-sensory inputs

In order to examine the effect of using a combination of irrelevant, redundant and relevant features extracted from the indoor data for occupancy monitoring, 50 different features from the sensors deployed were used in the analysis, refer to table (5.6). These features included irrelevant features such as those obtained from the VOC sensor, which was established in chapter four as poor predictors of occupancy numbers. Section 4.4.1 gives a detailed description of the feature relevance, redundancy and irrelevance analysis. Table (5.7) presents the model performance using features from all sensors as model input. MAPE was not used for this analysis because model produced using complete set of sensory features produced estimations that were very close to zero, and hence would not make sense to use.

Table (5.6): Complete set of sensory features used as model inputs

Climatic Data	Features notation					
	AVR	FDIFF	AF_DIFF	SDIFF	AS_DIFF	VAR
CO_2	✓	✓	✓	✓	✓	✓
VOC	✓	✓	✓	✓	✓	
Case temp	✓	✓	✓	✓	✓	✓

Sound Data	Features notation					
	TOTAL	FDIFF	AF_DIFF	SDIFF	AS_DIFF	VAR
TOS	✓			✓	✓	✓
TONP		✓		✓	✓	✓
THI_LO		✓	✓	✓	✓	✓
TLO_HI	✓			✓	✓	

PIR data	TOTAL	FDIFF	AF_DIFF	SDIFF	AS_DIFF	VAR
TOS	✓	✓	✓			✓
TONP		✓	✓	✓		✓
THI_LO	✓	✓	✓	✓	✓	✓
TLO_HI			✓			

Table (5.7): Performance metrics – Complete set of sensory features

Complete sensor features	RMSE	R^2	RAE
14th	1.850	0.373	0.614
17th	1.668	0.404	0.583
18th	1.509	0.479	0.522
19th	2.154	0.328	0.680
20th	1.602	0.427	0.565

Figure (5.17) compares the model performance when a selected optimal feature subset (FDIFF_CO_2, AF_DIFF_CO_2, AVR_CAS, VAR_CAS, VOS_SND and TOS_SND) was used, as opposed to the complete set of sensory features. Clearly, the estimation performance was quite poor compared with using the optimal multi-sensory features. The presence of redundant and irrelevant features adversely affects the classifier's performance, as they may have introduced complexity to the model's learning. In a typical week, an average daily RMSE of 1.757 was achieved using all sensory features, which is about an 85.80% increase compared to when an optimal feature subset was used. This is quite significant given the actual occupancy levels in the test area are between 0 and 6. There were also significant variations as well as relative error between model estimations and actual occupancy data with the use of the all sensory features inputs, as indicated with a relatively low average daily R^2 value of 0.402 and RAE of 0.593.

An occupancy detection system using complete set of sensory inputs, not only produces less reliable occupancy information, but also utilises measurements from redundant and irrelevant sensors which could increase instrumentation cost and computational time for occupancy monitoring, without any benefit to the system performance.

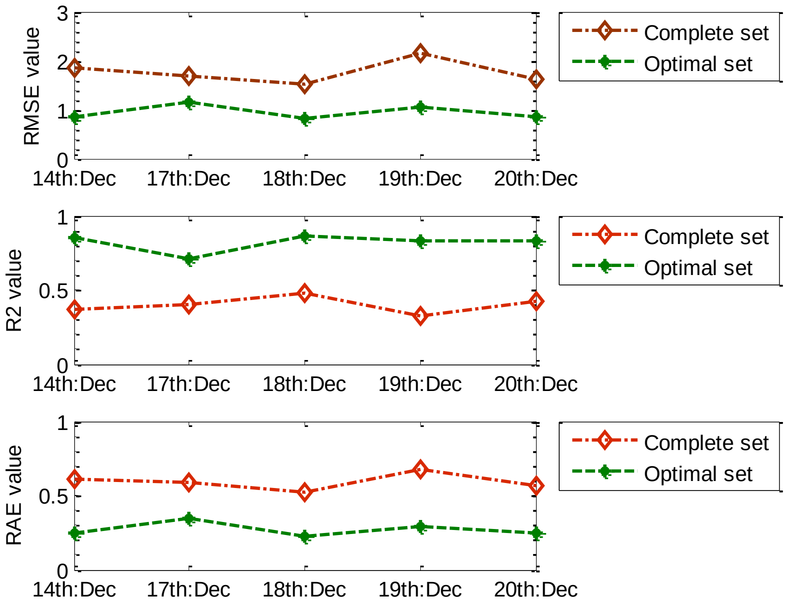


Figure (5.17): Comparison of optimal heterogeneous multi-sensory features and complete set of sensory features

5.7 Resilience in occupancy sensing network

In a multi-sensory approach for occupancy monitoring, several sensors often cooperate to provide real-time measurements. Due to shortcomings such as limited computational capacity, power consumption and band-width (Nakayama et al., 2007), sensors can temporarily become unreachable. Besides, their measurements may drift and lose calibration due to environmental interference. This may have ramification for an occupancy detection system using a multi-sensory sensor network.

In order to test the robustness of model estimations, selected features were excluded from the model inputs to stimulate sensor drop-out or failure. Table (5.8) shows the impact of case temperature sensors drop-off on the model estimations. Results were compared to those obtained from the heterogeneous sensor network described in section (5.4.5). Excluding case temperature features from fusion model inputs, reduced the model accuracy to 60.23%, which accounted for a 19.34% drop and increased the RMSE by 15.78% on the 18/12/2012. Variations between model estimations and actual occupancy data increased by 15.02% as indicated by the R^2 values while the relative error increased to 0.338. On the 19/12/2012, the impact of case temperature data exclusion is smaller compared to the previous day. RMSE value increased by 1.50%, R^2 value and accuracy decreased by 1.33% and 4.67% while the relative error increased by 5.03%. This trend is similar for 20/12/2012.

Table (5.8): Impact of case temperature sensors drop-off on the model performance

Days tested	RMSE	R^2	RAE	Accuracy
18th	0.943	0.730	0.338	60.23
19th	1.080	0.816	0.303	65.33
20th	0.877	0.832	0.292	67.27

Excluding sound features, the model RMSE value increased to 1.057 on the 18/12/2012, while the R^2 value dropped to 0.796. The relative error increased to 0.326, impacting on the model accuracy which dropped by 18.31%. On the 19/12/2012, removal of sound features, as with the two other scenarios (exclusion of CO_2 and case temperature features) showed a smaller impact on the model performance, with model accuracy reducing by 6.73%. This indicates a trend where the models in all three scenarios struggle to detect an additional occupant in the space. Table (5.9) shows impact of sound sensor drop-off on the model performance.

Table (5.9): Impact of sound sensor drop-off on the model performance

Days tested	RMSE	R^2	RAE	Accuracy
18^{th}	1.057	0.796	0.326	61.00
19^{th}	1.098	0.829	0.285	63.92
20^{th}	1.001	0.738	0.351	61.65

When CO_2 features were excluded from model inputs, the model accuracy was 59.32%, which amounted to a 20.56% drop on the 18/12/2012. This was slightly lower than when case temperature and sound features were excluded. Table (5.10) presents the impact of CO_2 sensor drop-off on the model performance. Over the three days tested, the CO_2 features drop accounted for the largest fall in model accuracy cumulating at an average of 14.73%. Sound and case temperature features drop-off saw the model accuracy fall by 13.09% and 10.08% respectively. This suggests that CO_2 measurements contributes signficantly to the model predictive ability, hence lend credence to the notion that a significant portion of CO_2 levels in the office is occupant related. The relatively large accuracy drop obtained without the use of sound features, coupled with their high predictive capacity as demonstrated in chapter four (section 4.4.6) may suggest that these sensors, although crude are capable of capturing occupancy related information.

It is clear that cooperation of sensors that provide strong occupancy information should be maintained always, if the model performance is to be kept stable. In the presence of possible individual sensor failure, control algorithms can be optimized to accommodate for this to ensure continuous reliable occupancy sensing.

Table (5.10): Impact of CO_2 sensor drop-off on the model performance

Days tested	RMSE	R^2	RAE	Accuracy
18^{th}	1.102	0.658	0.373	59.32
19^{th}	1.147	0.731	0.342	62.80
20^{th}	1.164	0.707	0.360	60.92

5.8 Alternative ventilation control sensors

While CO_2 sensors are commonly implemented as control sensors for ventilation in many buildings, limitations of CO_2-based occupancy sensing such as high cost and significant long term drift may form drivers for the development of alternative sensors for occupancy driven ventilation control.

A combination of low-cost case temperature and sound sensing can be considered as a promising alternative for use as control sensors in occupancy –based control. These sensors are cheaper in terms of actual product and installation cost. Also, they can also provide relatively good results for occupancy number estimation as indicated in section (5.6), where the model performance achieved an average accuracy of 61.01% for the days tested. Results suggest their applicability for occupancy numbers detection in a demand driven HVAC control strategy for an open plan office configuration, as they may serve as a far cheaper option than CO_2 sensors. However, more data from different building spaces may be required to establish this.

Temperature sensors used in this research cost about £40/unit, while the sound sensor can be assembled for just under £10/unit and even cheaper if mass produced. These sensors are easy to install in office buildings, with minimal retrofitting and disruption to occupants activities. Cost can be further reduced by using a small thermistor-based temperature sensor with the cheapest one costing less than £3; such as the iButton manufactured by Maxim integrated (Maxim integrated). For instance, deploying the iButton temperature sensors (x6) and custom sound sensor (x4) in the space may be a cost effective option if it can operationally replace the function of a single CO_2 sensor in the space.

5.9 Alternative approaches to occupancy estimation

To test the robustness of the proposed data processing methodology described in chapter four, several representative machine learning techniques were applied for occupancy estimation. The selected fusion techniques include SVM, radial basis neural network (RBF), and linear regression (LR). These techniques have been chosen at random, although they reflect what has been utilised for occupancy levels estimation in literature. Dong et al. (2010), Yang et al. (2012) and Mamidi et al. (2012) have applied SVM, RBF and LR respectively for their occupancy detection systems. RMSE, MAPE, RAE, R^2 alongside the Normalised root mean square error (NRSME) were used to evaluate the performance of all the above techniques. NRMSE is scale dependent and more suitable for comparing model estimations obtained using different algorithms, and has been applied to compare the performance of occupancy detection models (Erickson et al., 2009). This is often expressed in percentage as shown in equation (5.10). Lower values indicate better model performance.

$$NRMSE = \frac{RMSE}{Q_{o(MAX)} - Q_{o(MIN)}} \; x \; 100(\%) \tag{5.10}$$

Where $Q_{o(MAX)}$ is the maximum value of Q_O, and $Q_{o(MIN)}$ is the minimum value of Q_O.

5.9.1 SVM regression model

The quality of SVM models often depend on proper setting of SVM meta-parameters such as the R (regularization parameter), ε (insensitive loss function) and the chosen kernel function (Cherkassky and Ma, 2004). These parameters have to be carefully selected to ensure good performance generalization of the data. Most previous reported studies have applied the Gaussian radial basis function as the kernel function (Dong et al., 2010, Mamidi et al., 2012). In this thesis, this type of kernel function was implemented in the SVM model. There is no consensus among researchers on how to choose R and ε values. However, Cherkassy and Ma (2004) proposed that both parameters can be optimally computed based on equation (5.11) and (5.12). Both equations were adopted for the SVM Meta parameters computation in this thesis.

$$R = 3\sigma_y \tag{5.11}$$

Where σ_y is the standard deviation of the training outputs

$$\varepsilon = \tau\sigma\sqrt{\frac{\ln n}{n}} \tag{5.12}$$

Where σ the standard deviation of the noise level in the data, τ is a constant value and n is the total data instances. For SVM regression tasks, $\tau = 3$ gives a good performance for various data set sizes, noise levels and target functions , (Cherkassky and Ma, 2004). σ was estimated from the data using equation (5.13).

$$\sigma^2 = \frac{h}{h-1}\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \qquad (5.13)$$

Where h is a constant, and are typically small values (in the 2-6 range) corresponding to low-bias/high variance estimators (Cherkassky and Ma, 2004). h was assumed to be 6, since actual occupancy data show high temporal variability. y_i is the actual model data and $\hat{y}_i$ is the associated estimated data. After computation, an epsilon SVM regression module was implemented in *WEKA* with the parameters; R = 6.0282, and ε= 0.0483. Figure (5.18) shows the stages in the implementation of the SVM model

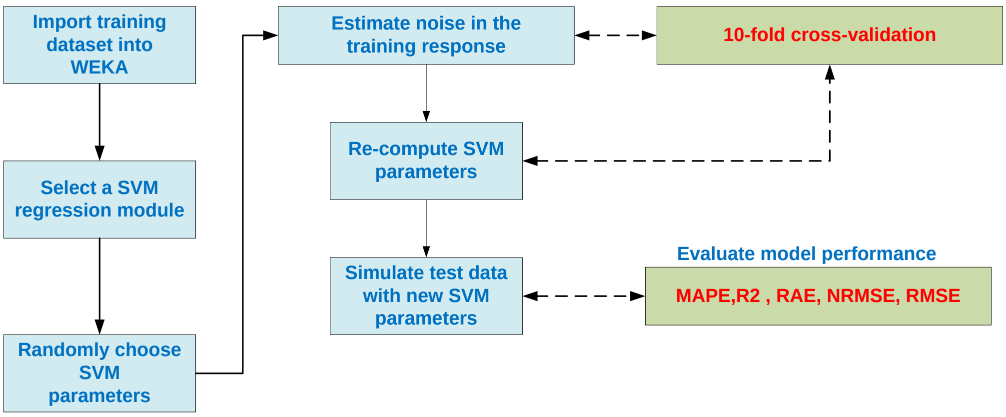


Figure (5.18) Stages in the implementation of the SVM model

Firstly, all data were used as the training set with randomly chosen values for R and ε . Noise levels (σ) in the data was estimated data using equation (5.13). This information was then used to compute new values for R and ε, before simulation of the test data was carried out.

5.9.2 Linear regression model

This model uses the Akaike information criterion (*AIC*) for model selection. *AIC* selects a model from a set of models, by choosing the one that minimizes the

Kullback-Leibler distance between the model estimations and actual data (Burnham and Anderson, 2002). This is based on an information theoretical heuristic, which seeks to find a model that closely fits ground truth data. *AIC* is given by equation (5.13).

$$AIC = 2k - 2\ln(L) \tag{5.14}$$

Where k is the number of the number of free parameters in the model and L is the maximised value of the likelihood function for the estimated model. A detailed description of *AIC* can be seen in the work of Burnham and Anderson (2002). In this thesis, the linear regression module used a ridge regularizer $= 1.0^{e-8}$ and an M5 attribute selection. The LR model was implemented in *WEKA*. The trained model estimates the number of occupants based on a linear combination of the input feature values.

5.9.3 Radial basis function model

In this work, the RBF regressor model was implemented in *WEKA*. It uses the Broyden–Fletcher–Goldfarb–Shanno (BFGS) optimization algorithm for minimizing the squared error between model estimations and actual data. BFGS algorithm is used for solving unconstrained non-linear optimization problems and details of this can be found in Byrd and Nocedal, (1989), and Byrd et al. (1987).

5.9.4 Comparison of models

With respect to all metrics applied in the analysis, all machine learning models showed similar trend on any particular day. Figure (5.19) and (5.20) show the occupancy estimations using four different models for two typical days in a week. The NRMSE for the NN, LR, SVM and RBF showed error rate of 13.58%, 15.68%, 14.10% and 14.82% and the corresponding RMSE values 0.815, 0.941, 0.846 and 0.869 respectively on the 18/12/2010 (refer figure to 5.21 and 5.22).

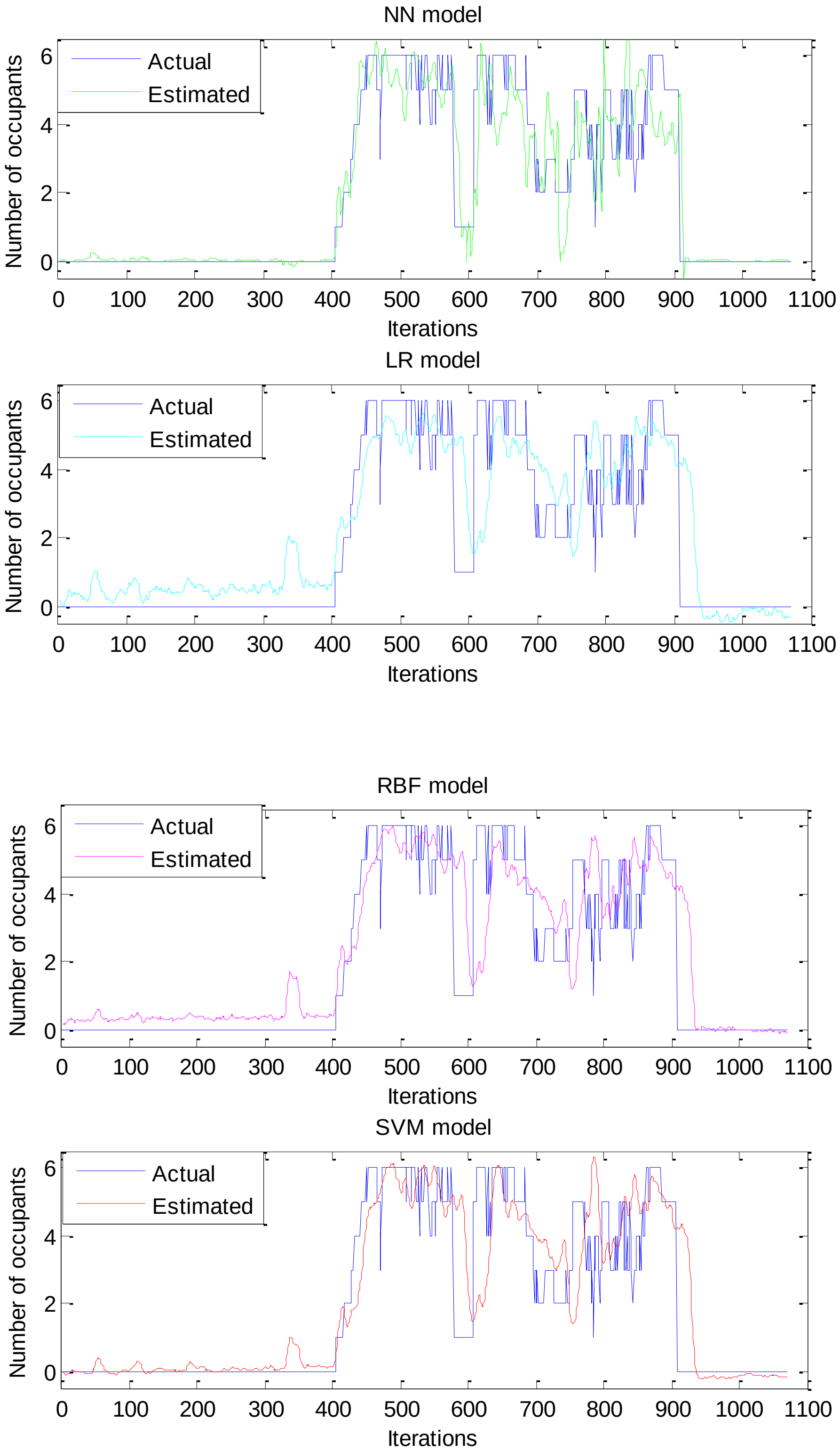


Figure (5.19): Comparison of different machine learning models on the 18/12/2012

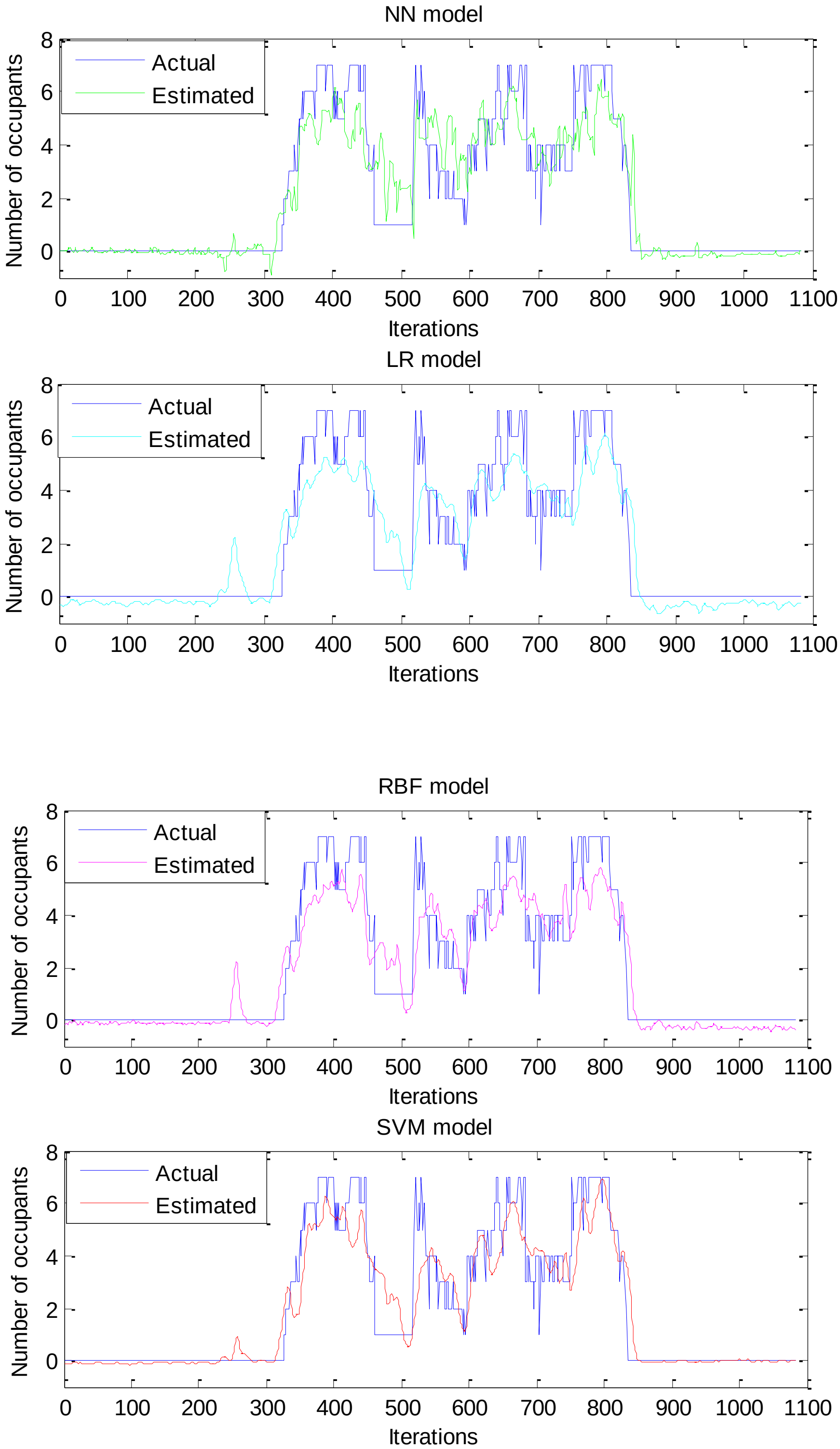


Figure (5.20): Comparison of different machine learning models on the 19/12/2012

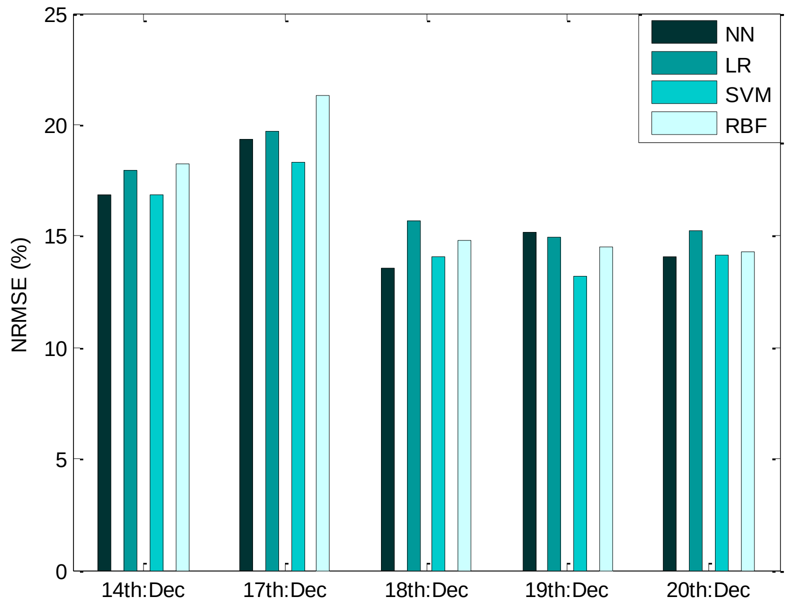


Figure (5.21): Comparison of NRMSE of all four models

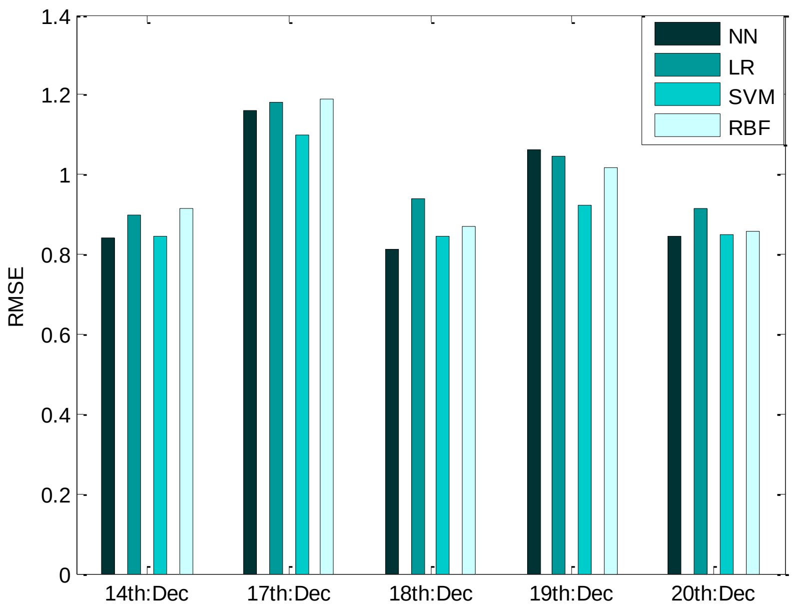


Figure (5.22): Comparison of RMSE of all four models

SVM, LR and RBF generated rather noisy estimates with frequent variation as against the true occupancy number, especially during the early hours of the morning. This may be so in part, because both SVM and RBF assume each data instance is independent and identically distributed. This may not always be the case, as parameters such as CO_2 levels tend to have strong inherent temporal correlations. It is also clear that while LR model can produce decent results, it does not achieve optimal occupancy estimations. However, it does provide some information about the model. For example, on the 18/12/2012, the LR model suggests that case temperature and CO_2 features are highly correlated to the number of occupants. Surprisingly, sound features produced poor correlation. The model learns the following coefficients for estimating observed number of occupants. Equation (5.10) shows this relationship.

$$\hat{O} = 0.2935 * AVR\_CAS - 0.2506 * VAR\_CAS + 3.2807 * FDIFF\_CO_2 - 2.3487 * AF\_DIFF\_CO_2 + 0.0069 * TOS\_SND + 0.068 * VOS\_SND - 5.5864 \qquad (5.15)$$

Where $\hat{O}$ is estimated occupancy number.

On the 19/12/2012, SVM had the least error rate at 13.22%, followed by RBF and LR at 14.52% and 14.94% respectively. NN showed the worst performance with an error rate of 15.20%, showing large fluctuations during occupied times. As already mentioned in this section, the rather noisy estimates may be linked to the assumption that the NN model also treated each data instance as independent and identically distributed.

The average accuracies achieved for NN, SVM, LR and RBF in the week tested were 70.15%, 71.34%, 67.25% and 68.22% respectively, with SVM providing the best accuracy, figure (5.23). All four models struggle to track abrupt changes in occupancy numbers within a short duration. This may be linked to the slow response of the CO_2 sensors, whose features contributes significantly to the model predictive capacity. However, for occupancy driven HVAC control, results from all four models may be considered sufficient, since sudden occupancy level changes may not be of any control significance, unless the space is switching from occupancy to vacancy or vice-versa.

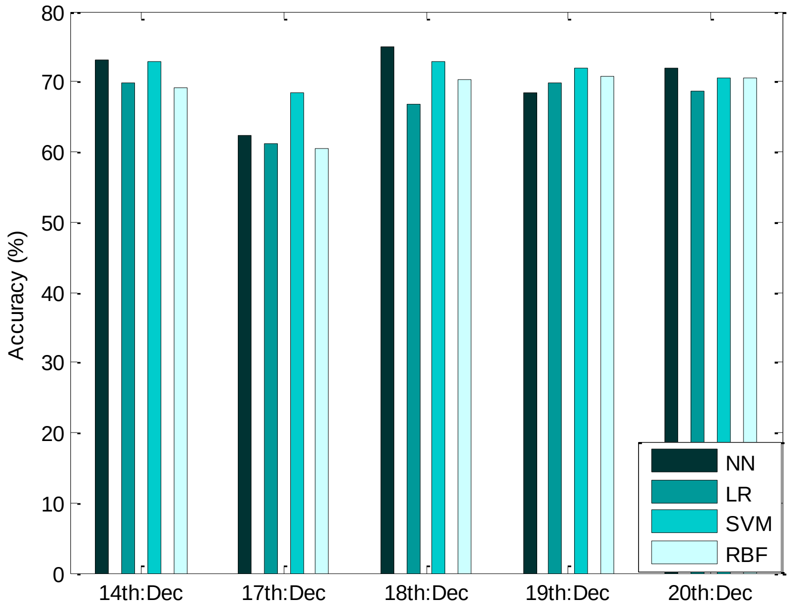


Figure (5.23): Comparison of occupied times accuracies for all four models

All models do show certain level of fluctuations and spikes in their estimation profiles as against actual occupancy data. This may be attributed to the models use of features inputs (such as FDIFF_CO_2, AF_DIFF_CO_2), which may be sensitive to noise. The average daily R^2 values indicated that LR model showed the largest variation with a value of 0.775, and SVM model showed the least with a value of 0.822. NN and RBF were 0.815 and 0.795 respectively. Results for the relative error show similar pattern. Figure (5.24) shows how the R^2 values vary for all four models in a typical week, while figure (5.25) presents that of the RAE values.

Extensive fluctuations in the estimated occupancy profile is highly undesirable for building operations, as this would lead to frequent adjustment of HVAC equipment, which may impact negatively on the equipment life-cycle. Hence, it could be useful to design appropriate filters to reject outliers in the estimated results, before integrating such outputs into any existing BEMS.

In summary, all the four machine learning techniques are seen to be showing the same tendency. Their model estimations show good tracking with actual occupancy numbers, and the SVM model performed better than others. This analysis confirms the robustness of the occupancy detection methodology applied in this thesis. Model results suggest the potential for generating occupancy information for enhancing building control operations, especially demand –driven HVAC systems.

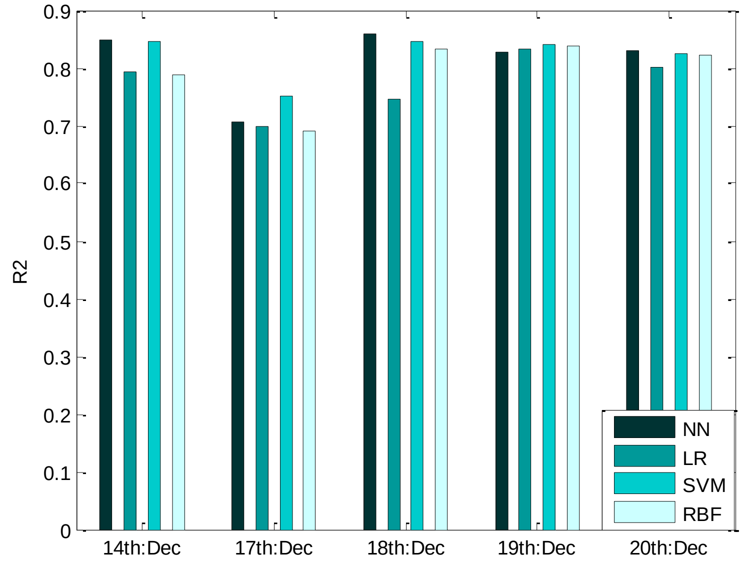


Figure (5.24): Comparison of R^2 of all four models

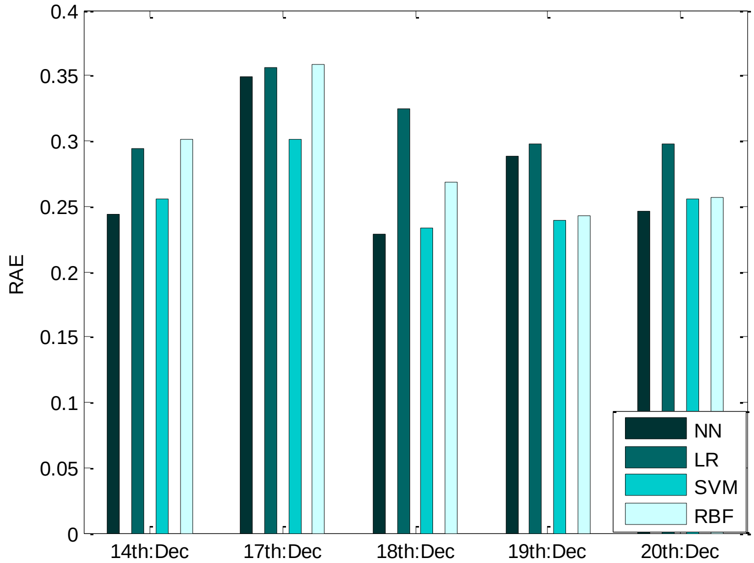


Figure (5.25): Comparison of RAE of all four models

5.10 Cross room model analysis

In order to examine the generalisation ability of a learned model for occupancy levels estimation in different spaces, a cross room model analysis was carried out. In this analysis, models trained with data from test area two and three were cross applied for occupancy levels estimations in both areas. For instance, the optimal features for test area two as shown in table (5.11) were extracted from test area three data, and then utilised to train an NN model before being applied for occupancy level estimations in test area two. The same process was repeated in the cross room model analysis for test area three. Same NN configuration described in section (5.3) was implemented as the fusion model. Sensor data were collected for seven days between 08/07/2012 00:00:00am and 16/07/2012 10:45:00am. For each daily result presented, the NN model was trained using six days' worth of data, while the remaining day was used for model testing.

Table (5.11): Optimal features for both test areas

Test area two model inputs	Test area three inputs
TOS_SND	FDIFF_CO_2
THI_LO_SND	AF_DIFF_CO_2
TOS_PIR	AVR_CAS
	VAR_CAS
	VOS_SND
	TOS_SND

5.10.1 Cross room model estimation: Test area two

In test area two, as expected regular patterns in occupancy data are observable for the two typical week days shown in figure (5.26) and (5.27). Occupancy levels were low in the early hours of the morning till about 8:00am and steadily rose, peaking during lunch time between 12:00 and 4:00pm, when both staff and PhD students (including occupants resuming late) would have arrived in most cases, if they are in for the day. Occupancy levels decreased again, as occupants finish for the day and vacate the space. Some researchers normally stay long into early hours of the next morning. Hence, results showed marginal occupancy levels during this period. Model estimations clearly track well with occupancy numbers, with accuracy, RMSE, R^2, RAE of 72.88%, 1.587, 0.818 and 0.251 respectively on the 10/07/2012. While on the 11/07/2012, accuracy, RMSE, R^2, RAE of 70.29%, 1.638, 0.807 and 0.260 was achieved. These results are considered good since the number of occupants in this test area varied between 0 and 11. This suggests that the model estimations were usually within 2 of the actual occupancy data. However, estimations obtained by applying a model trained using data from test area three produced poor results with an RMSE, R^2 and RAE of 3.375, 0.253 and 1.254 respectively on the 10/07/2012, while on the 11/07/2012 an RMSE, R^2 and RAE of 2.884, 0.330, and 1.158 respectively was obtained, as shown in table (5.12).

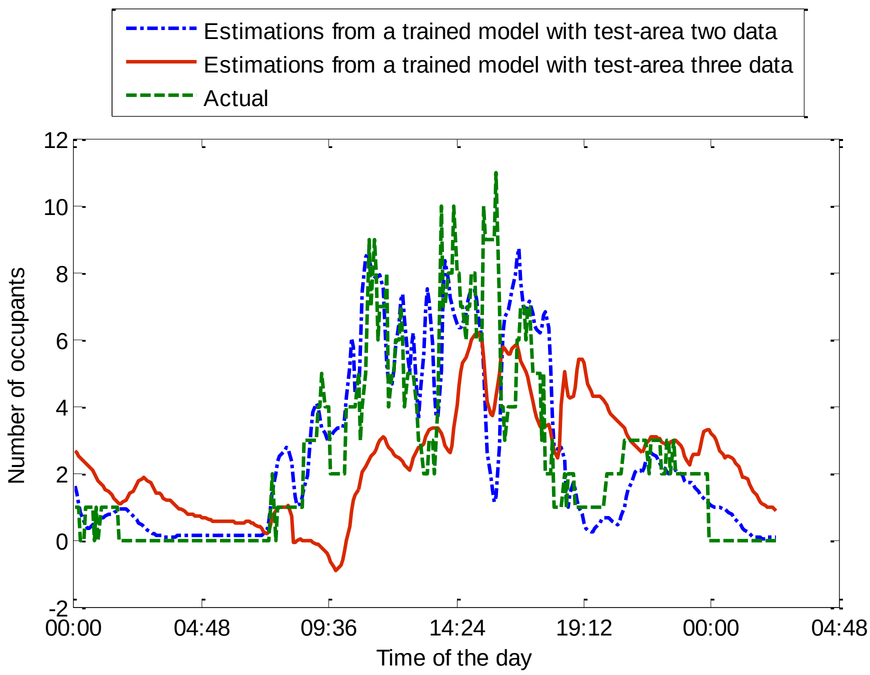


Figure (5.26): Occupancy estimations for 10/07/2012- Test area two

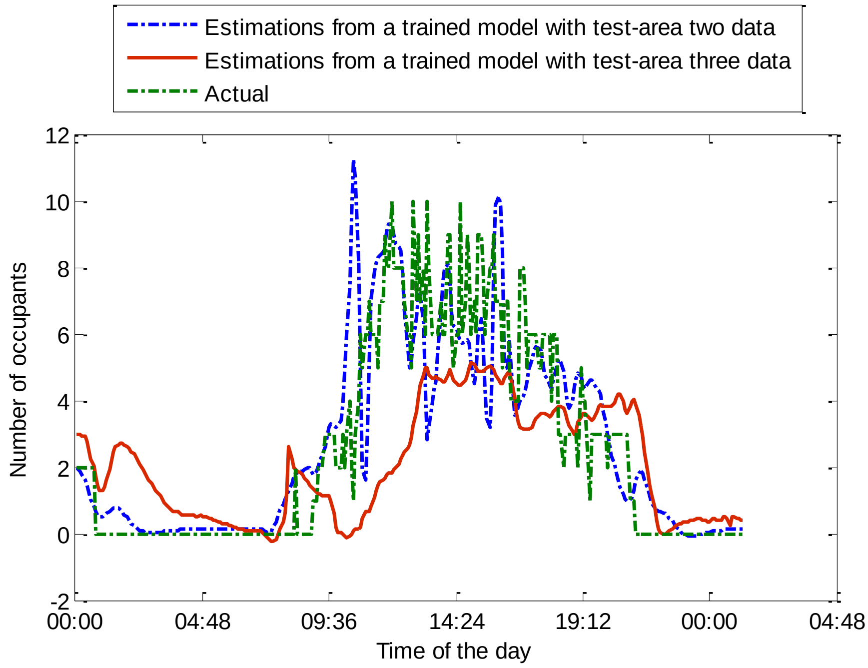


Figure (5.27): Occupancy estimations for 11/07/2012- Test area two

Table (5.12): Cross room model analysis for test area two

Days	Model trained with test area two Data				Model trained with test area three data		
	RMSE	R^2	RAE	Accuracy(%)	RMSE	R^2	RAE
10^{th}	1.587	0.818	0.251	72.88	3.375	0.253	1.254
11^{th}	1.638	0.807	0.260	70.29	2.884	0.330	1.158

5.10.2 Cross room model estimation: Test area three

In test area three, estimations from a trained model using data from this test area were in good sync with the actual occupancy data as shown in figure (5.28) and (5.29). Model estimations track actual occupancy levels with an accuracy, RMSE, R^2 and RAE of 63.93%, 1.1049, 0.726 and 0.328 respectively on the 10/07/2012, while on the 11/07/2012 with an accuracy, RMSE, R^2 and RAE of 81.17%, 0.782, 0.882, and 0.210 respectively was obtained, as shown in table (5.13).

From figure (5.28) and (5.29), model trained from test area two data clearly tend to overestimate occupancy numbers in test area three. This may be so because of higher entropy of indoor variables in test area two due to higher actual occupancy. Table (5.13) shows the results of the cross room analysis. Model estimations failed to show any reasonable tracking with actual occupancy levels. RMSE and RAE of 3.930 and 1.595 were obtained on the 10/07/2012, while that of on the 11/07/2012 were 3.268 and 3.934. R^2 values achieved on both days tested suggested that there were significant fluctuations between actual occupancy data and model estimations.

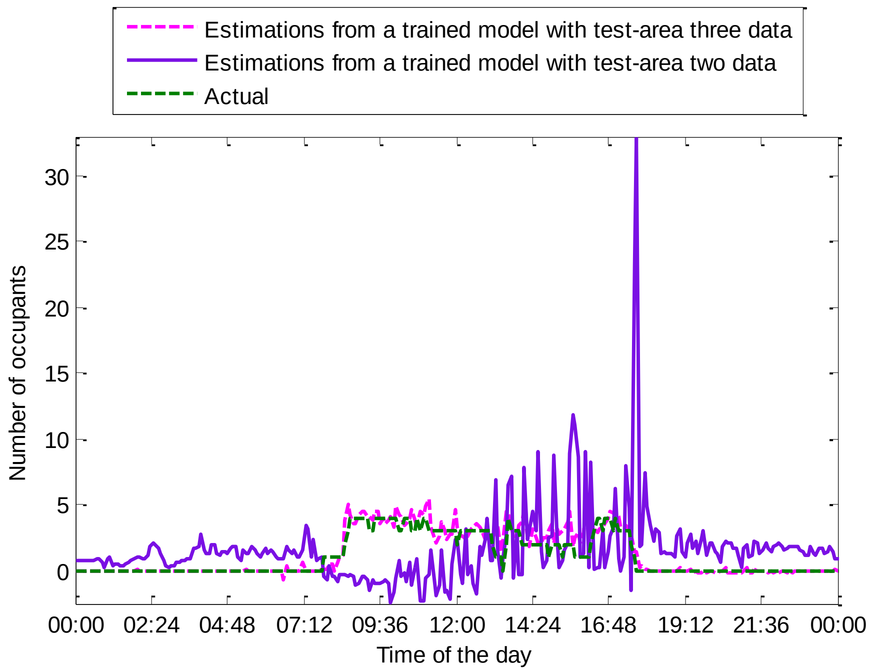


Figure (5.28): Occupancy estimation for 10/07/2012- Test area three

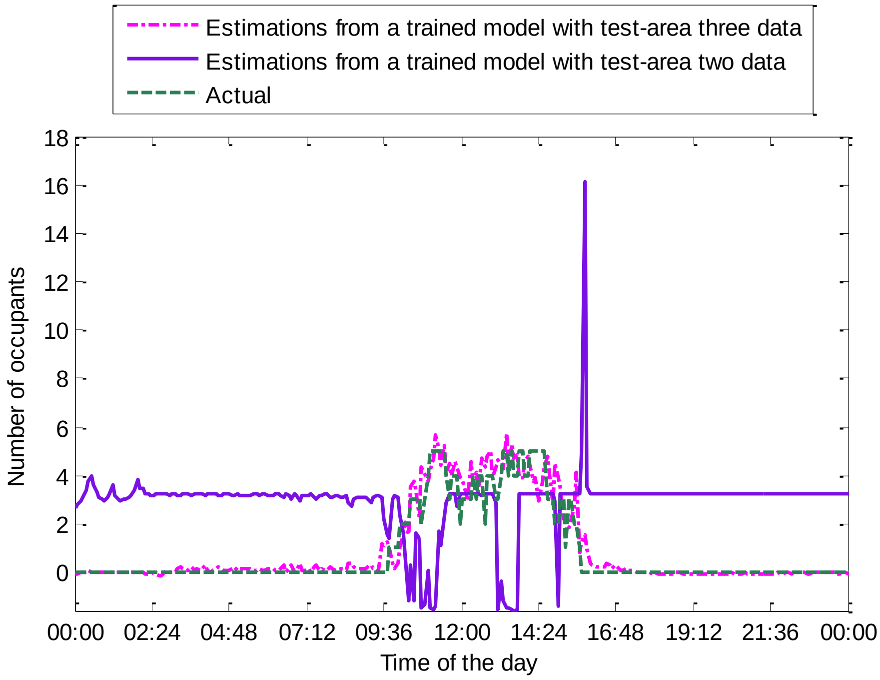


Figure (5.29): Occupancy estimation for 11/07/2012- Test area three

Table (5.13): Cross room model analysis for test area three

Days	Model trained with test area three Data				Model trained with test area two data		
	RMSE	R^2	RAE	Accuracy(%)	RMSE	R^2	RAE
10th	1.1049	0.726	0.328	63.93	3.930	-0.437	1.595
11th	0.7822	0.882	0.210	81.17	3.268	-3.744	3.934

The poor results obtained from the cross room model analysis were not surprising, as the indoor environmental dynamics in both rooms are different. For instance, CO_2 decay rates are much slower in test area three than that of test area two, where the former is generally tighter, with less openings compared to that of the later. Besides, actual occupancy levels in test area three was lower than that of test area two, and as such may have affected the indoor environmental dynamics. This may have accounted for the poor learning when applied in cross room model training, and thus resulted in significant deviations from actual occupancy data in the model estimations. Future work can include ways of standardizing the model performance for generalization in similar spaces within a building.

5.11 Weekend estimations

For weekends, accuracy was obtained by dividing the number of correctly estimated data instances by the total number of data instances, as MAPE is limited due to zeroes in the data. Researchers work during weekends in test area two, and there are relatively less changes in indoor environmental variables, as opposed to that during working days. Model estimations were impressive, and track actual occupancy levels with an accuracy of 68.11% for 14/07/2012, and accuracy of 84.59% on the 15/07/2012, as shown in figure (5.30) and (5.31).

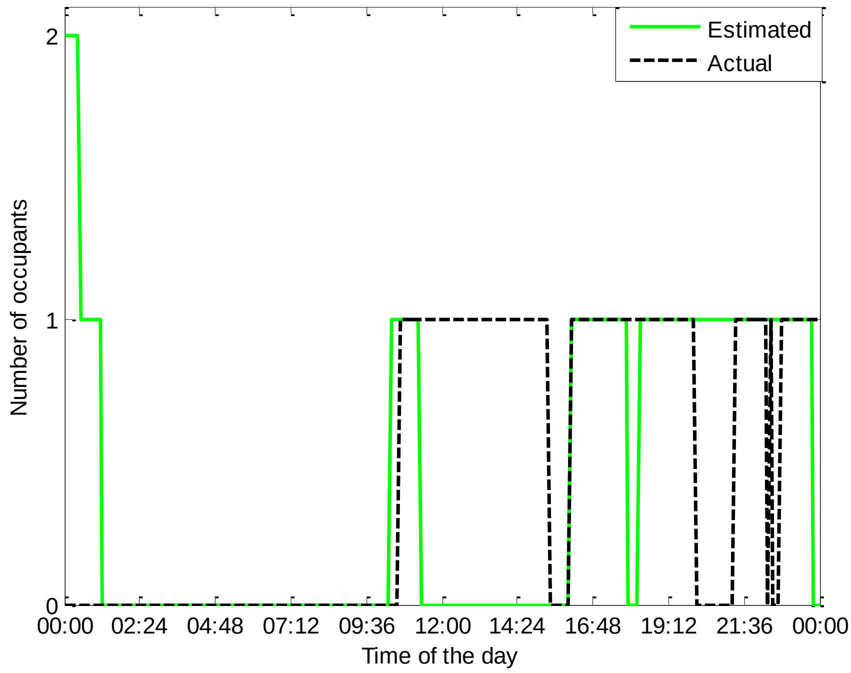


Figure (5.30): Occupancy estimation for 14/07/2012- Test area two

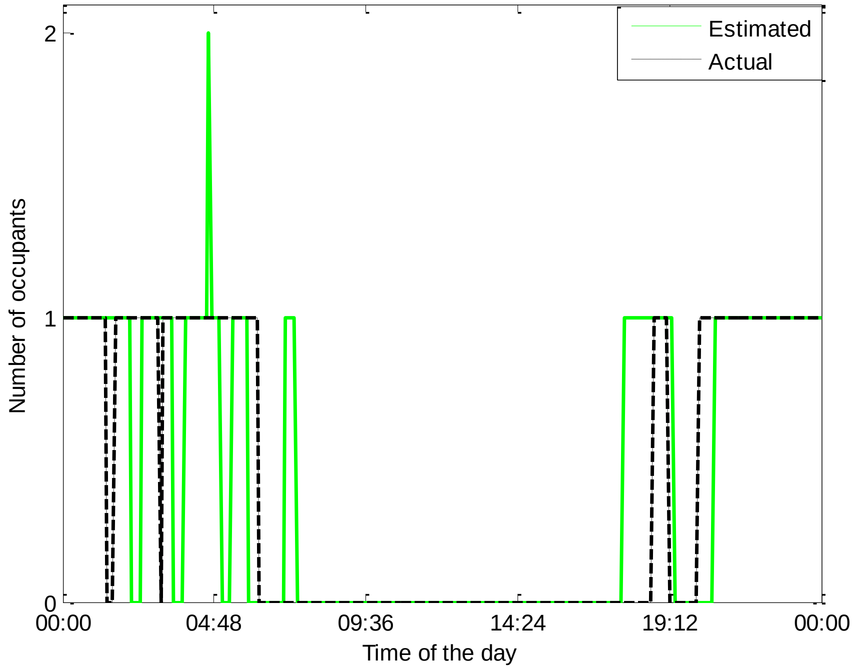


Figure (5.31): Occupancy estimation for 15/07/2012- Test area two

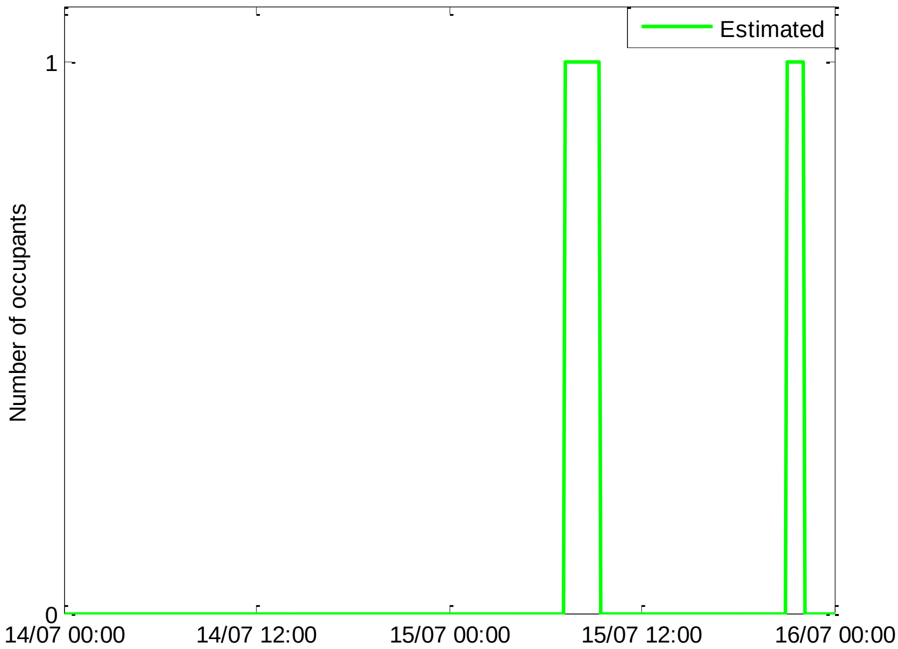


Figure (5.32): Weekend occupancy estimation –Test area three

In test area three, actual weekend occupancy data was zero during the data collection periods, hence was not shown in figure (5.32). Model estimations were obviously better, as indoor climatic conditions in the space were fairly constant. On the 14/07/2012 estimations accuracy was 100%, while on the 15/07/2012 it was 87%. These results may suggest the model efficacy for detecting space occupancy and vacancy, which may be crucial in any occupancy- driven HVAC operation.

5.12 Chapter summary

This chapter presents occupancy estimation results obtained using different sensory configurations, comprising of low-cost and non-invasive sensors. A number of performance metrics have been applied for testing and validating of the sensor networks, with results suggesting that a heterogeneous multi-sensor network for occupancy estimation provides more reliable results compared to the use of redundant sensor network. The heterogeneous multisensory network comprises of selected sensors (a set of crude indoor environmental sensors) capable of capturing the maximum occupancy information in an open- plan office. The use of alternative low cost sensors (such as case temperature and sound sensors) to assist in ventilation control strategies, instead of CO_2 based sensors can have compelling implications on the building energy use and air quality control.

The concept of resilience in an occupancy sensing network was highlighted, with CO_2 sensor drop-off causing a larger reduction of the model accuracy (reaching an average of 14.73%) than sound and case temperature sensor drop-off. Various data driven machine learning models such as SVM, LR and RBF have been applied to test the robustness of the data processing methodology proposed in this research, with SVM model producing the best estimations. In general, the model estimations tracked actual occupancy levels with accuracy reaching 81.17% during week days. However, the model produces poor results when applied for occupancy estimation in a space different from where its training data was obtained.

CHAPTER 6

MINING INDOOR ENVIRONMENTAL DATA FOR EFFECTIVE BUILDING MONITORING AND ENERGY EFFICIENCY

6.1 Introduction

Any useful knowledge on the relationship between various indoor environmental variables that can be extracted from monitoring data may be beneficial to the enhancement of building energy management. In this chapter, section 6.2 explores the relationship between VOC and CO_2 levels and its implication for ventilation control, while section 6.3 looks into the relationship between electricity use, average case temperature and occupancy levels. Section 6.4 demonstrates the potential for energy savings through occupancy driven ventilation control, and finally section 6.5 presents the chapter summary.

6.2 Ventilation control: VOC or CO_2 sensing

The time frame used for this analysis covered a period of three weeks between 27/11/2012 and 20/12/2012, which provides sufficient time for daily temporal trends to be observed. As mentioned earlier in the thesis (section 2.4.6), there is a growing debate on the most effective demand controlled ventilation strategy to implement in buildings; VOC or CO_2 or both. A clear understanding of the relationship between both parameters in an observed space may be useful information for the development of an effective and efficient IAQ monitoring strategy. For instance, CO_2 decay patterns in buildings with high occupancy may be useful to determine air change rate, and occupancy patterns, and how a variable air volume (VAV) system may be modulated (Aglan, 2003). A building that takes longer time for CO_2 to decay may suggest it is tighter than that which takes shorter time. While this may be a plus for ensuring energy efficient building services, it also presents a challenge for maintaining acceptable IAQ.

From figure (6.1) and (6.2), VOC and CO_2 levels show a similar trend, with both peaking at similar times. A regression analysis confirms this pattern with a strong R^2

value of around 0.70, see figure (6.3). This is quite surprising, given that the observed environment is a pure office setting and both parameters are rather independent of each other. This seems to be contrary with findings in previous research, where both parameters show poor or no correlation such as in the work carried out by Painter et al. (2012). The implication of this may suggest that both parameters can be used for common functions or may even replace one another in a DCV strategy. There is a growing trend to install VOC sensors for IAQ monitoring due to its cheaper installation cost compared to CO_2 sensors. However, while CO_2 levels were clearly linked to space occupancy (Emmerich and Persily, 2001), hence their high predictive capacity as demonstrated in section (4.5.4) and supported by previous research (Emmerich and Persily, 2001), VOC levels on the other hand show poor correlation with occupancy numbers.

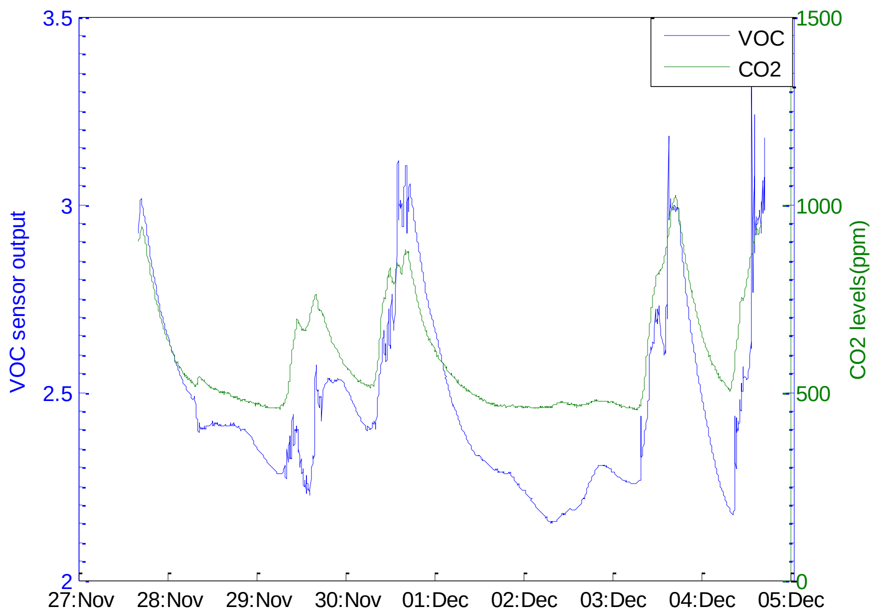


Figure (6.1): Air quality measurements between 27/ 11/2012 and 05/12/2012

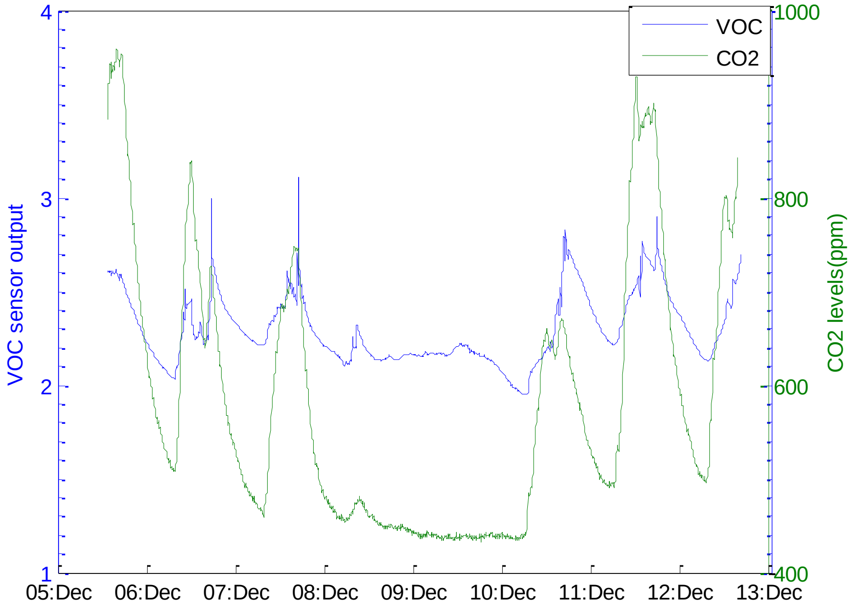


Figure (6.2): Air quality measurements between 05/ 11/2012 and 13/12/2012

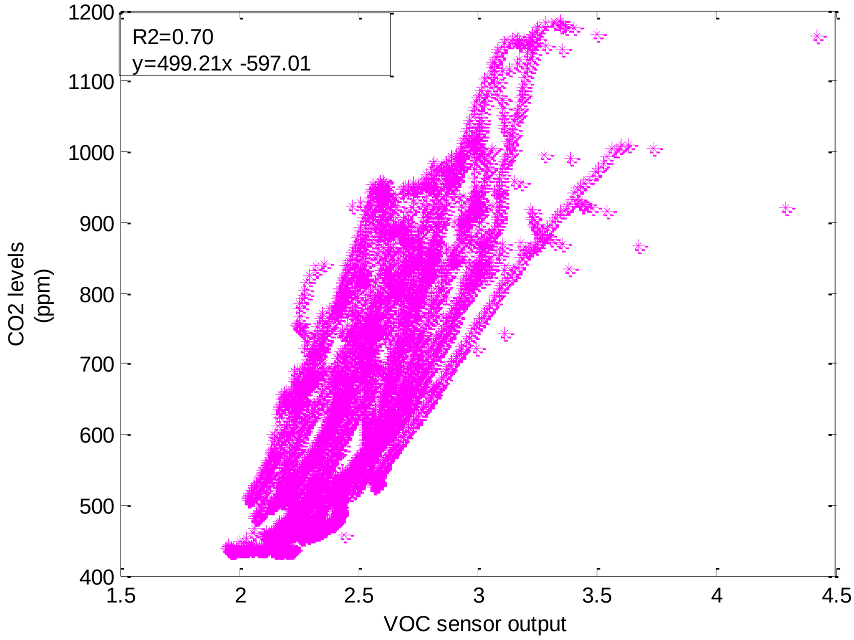


Figure (6.3): VOC and CO_2 scatter plot between 27/ 11/2012 and 20/12/2012

VOC emissions have been linked to the usage patterns of personal computers in offices (Berrios et al., 2003, Funaki et al., 2003). It was reported that VOC levels were higher by 10-120 times when the computers were ON than when they are OFF (Berrios et al., 2003). From figure (6.4), the average case temperature trace clearly tracks VOC levels, and may have been influenced by office equipment usage patterns in the space. Case temperature measurements reflect the use of office equipment during occupancy. Again, results from a linear regression analysis between case temperature and VOC measurements does support this, with R^2 values obtained reaching 0.87 in some days, see figure (6.5) – (6.6). During the weekend, where indoor climatic parameters are fairly constant, both parameters produced an R^2 value of 0.63 as shown in figure (6.7). Variations in R^2 values for different days may be due to outdoor temperature and relative humidity variations, which are known to affect indoor VOC levels in buildings (Wolkoff and Kjærgaard, 2007). However, such VOC levels fluctuations may have ramifications for DCV systems, as this may cause over-ventilation in some days, which could increase energy use or under-ventilation, thereby comprising IAQ. In addition, as mentioned previously in section (2.4.6), outputs from VOC sensors are not calibrated for specific pollutants, which make them difficult to use. While results from the observed space (test area three) are promising, further work is recommended to establish VOC sensors as a possible replacement for CO_2 sensors in DCV applications.

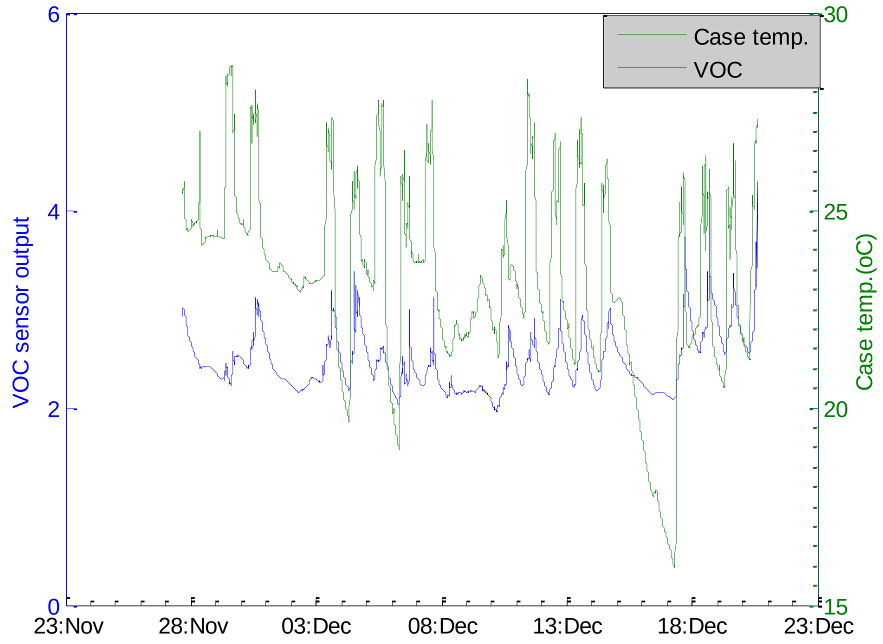


Figure (6.4): Average case temperature and VOC levels

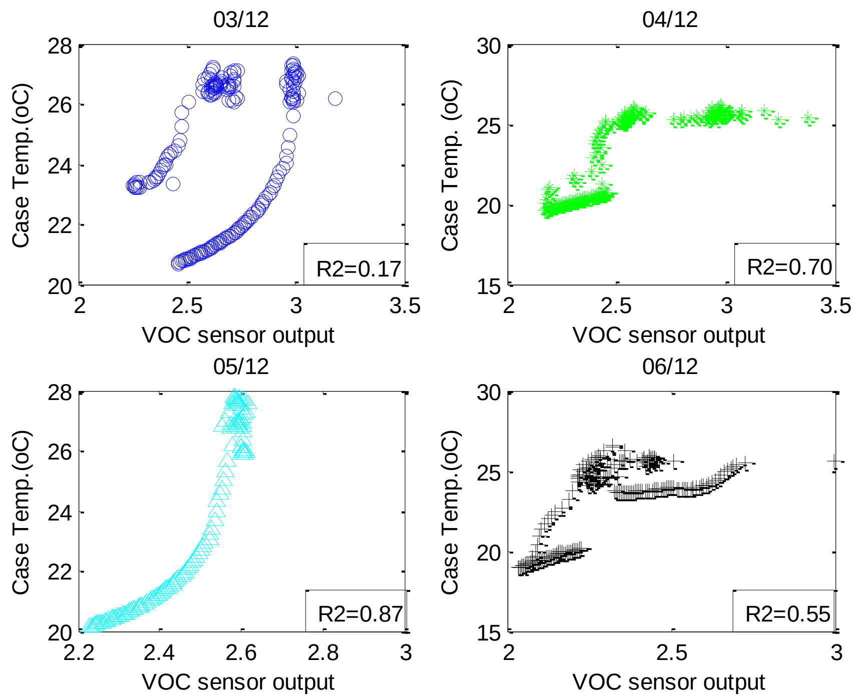


Figure (6.5): Case temperature and VOC regression 03/12-06/12

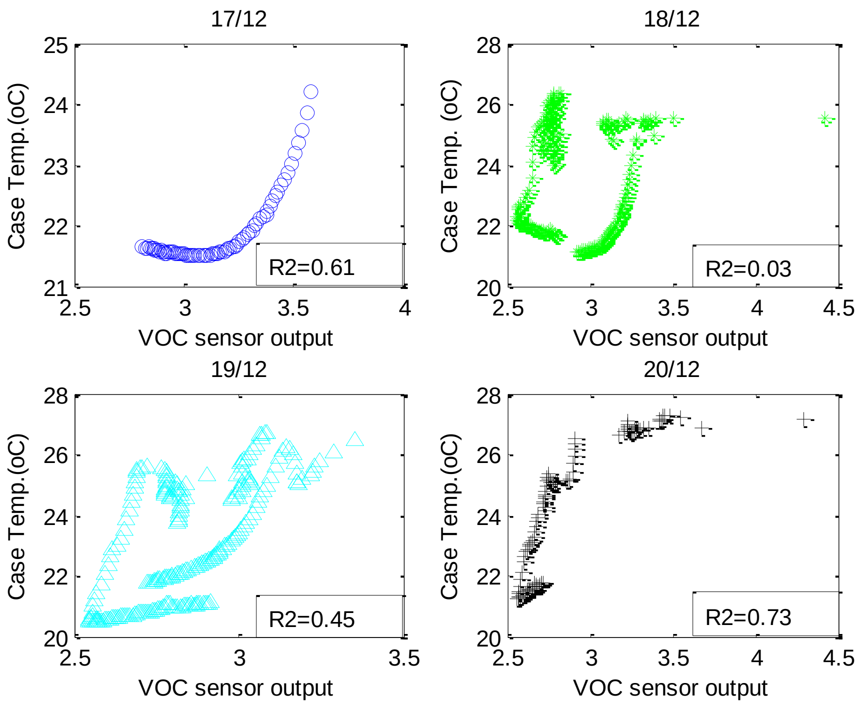


Figure (6.6): Case temperature and VOC regression 17/12-20/12

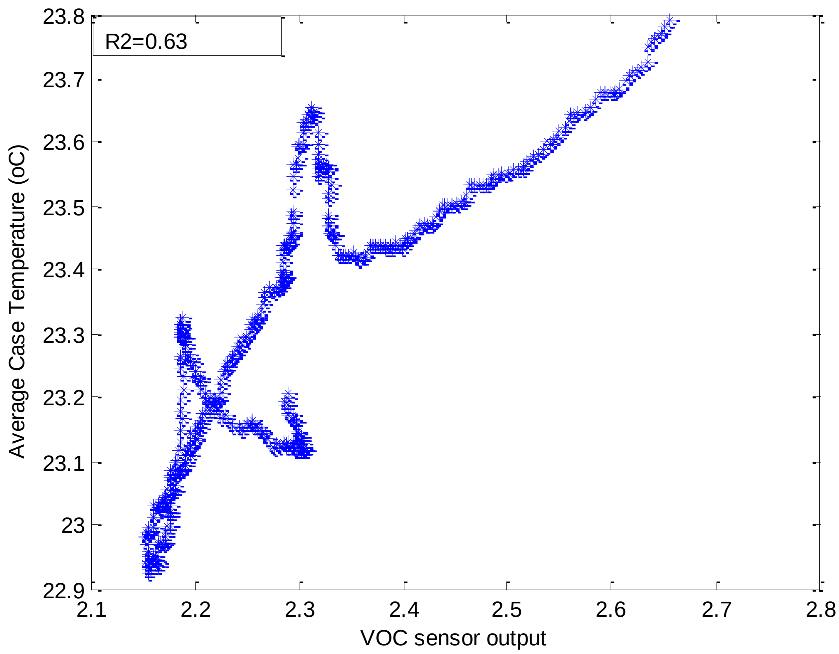


Figure (6.7): Case temperature and VOC regression 01/12-02/12 (Weekend)

6.3 Electricity use, case temperature measurements and occupancy levels

Electrical appliances can use up to 40% of office building electricity (IEA, 1997), and as such office equipment usage can be considered as a key variable in energy use studies in many buildings. Hence, the relationship between energy (electricity) and equipment usage is of particular interest. Regression analysis was used to study the data. Figure (6.8) shows a plot of the average case temperature (from which usage patterns can be inferred) and electricity use over a period of about four weeks. Both parameters fall and rise based on space occupancy or vacancy. A typical week in the monitoring period was further studied to examine any relationship between both parameters. For the days examined as shown in figure (6.9), the highest R^2 value of 0.58 for the relationship between average case temperature and electricity use was achieved on the 19/12/2012, where 7 persons used the space instead of the 6 persons normally using the space most of the time. Other days recorded R^2 values of 0.40, 0.42, and 0.32 on the 17/12/2012, 18/12/2012 and 20/12/2012.

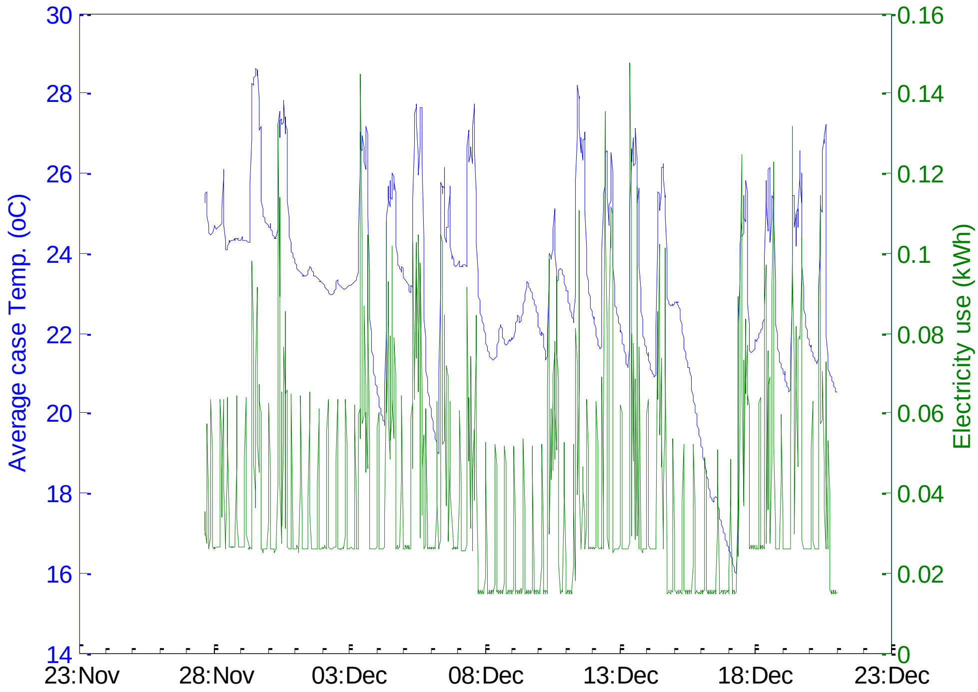


Figure (6.8): Average case temperature and electricity use

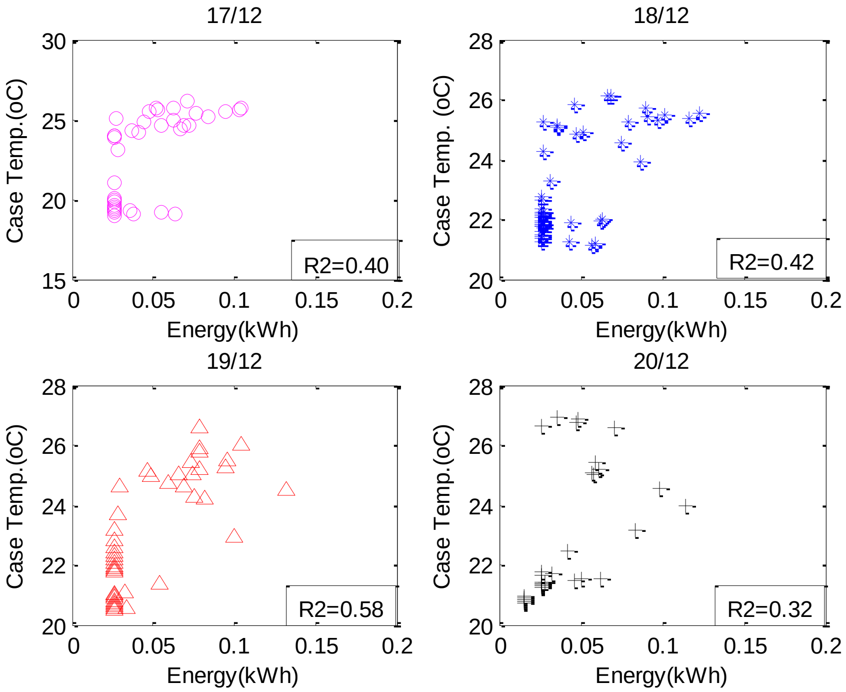


Figure (6.9): Average case temperature and energy (electricity use) regression 17/12-20/12

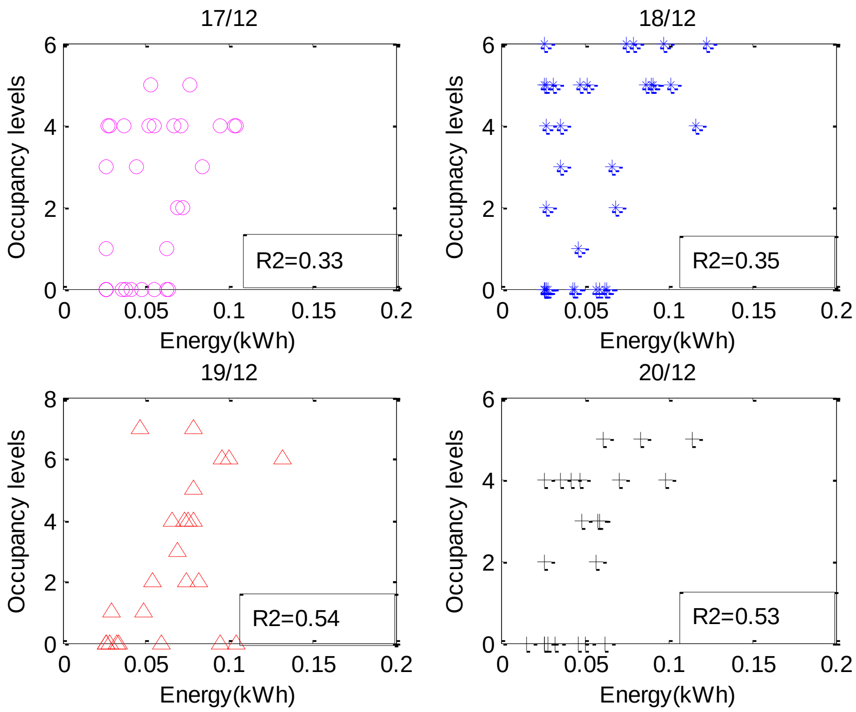


Figure (6.10): Occupancy levels and energy (electricity use) regression 17/12-20/12

In figure (6.10), the relationship between occupancy levels and electricity use were similar on the 19/12/2012 and 20/12/2012, with R^2 values of 0.54 and 0.53 respectively. A value of 0.33 was achieved on the 17/12/2012, and 0.35 on the 18/12/2012. This suggests that occupancy may be a useful predictor of energy use. However, there was no clear pattern between both regression studies. Variations in the analysis may be so because the sub-meters monitoring the space covers other electrical circuits (such as printer, kettle, and lighting), and in addition to the desktop computers circuits instrumented for case temperature.

Figures (6.11) – (6.14) presents the entropy in electricity use, average case temperature and occupancy levels. Changes in electricity use and average case temperature show fairly similar pattern, with the exception on the 20/12/2012. This may be supported by the previous evidence of its relatively poor R^2 value of 0.32 achieved on this day, refer to figure (6.9). Change in occupancy levels and energy use were in good sync on the days examined, especially on the 19/12/2012. However, during unoccupied periods spikes in the electricity use data may be due to lightings that were left ON at the close of the previous day's work.

In summary, there is some relationship between occupancy, energy (electricity) use and case temperature, although it is difficult to clearly define. A clearly described relationship between the variables may be beneficial to the implementation of more effective power management strategies for electronic appliances in office buildings. For instance, a clear link between electricity use and case temperature measurements may give insights about the switch-off rates of appliances such as desktop computers. This can in turn provide information on what time of the day and location within an office building where possible energy savings could be more significant.

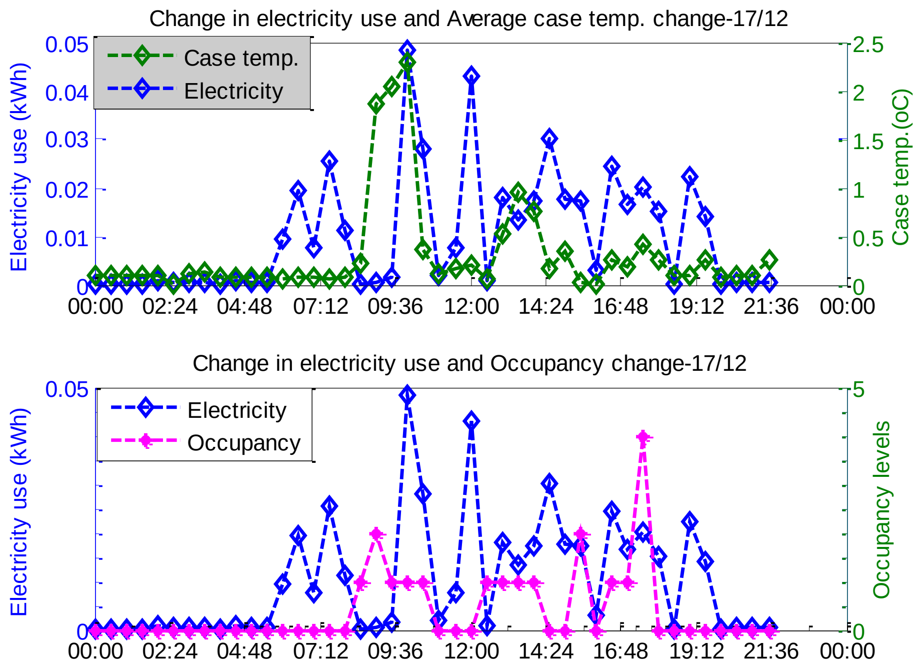


Figure (6.11): Change in electricity use, average case temperature, and occupancy levels-17/12

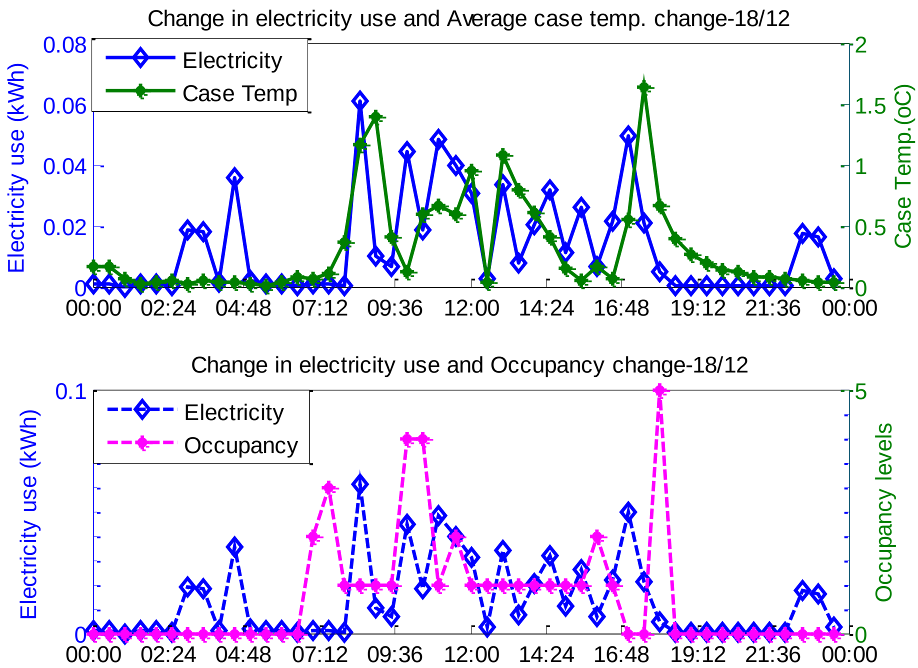


Figure (6.12): Change in electricity use, average case temperature, and occupancy levels-18/12

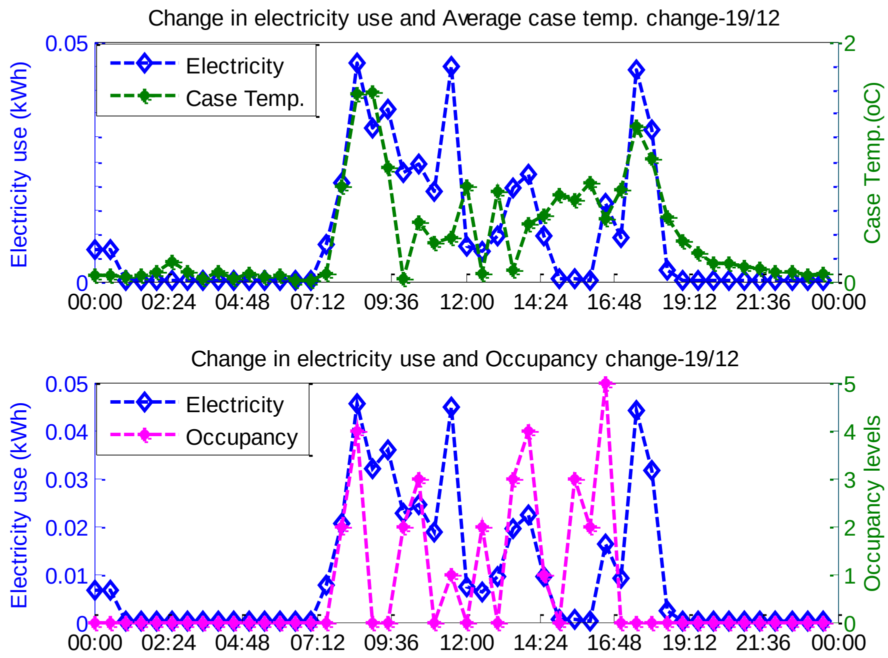


Figure (6.13): Change in electricity use, average case temperature, and occupancy levels-19/12

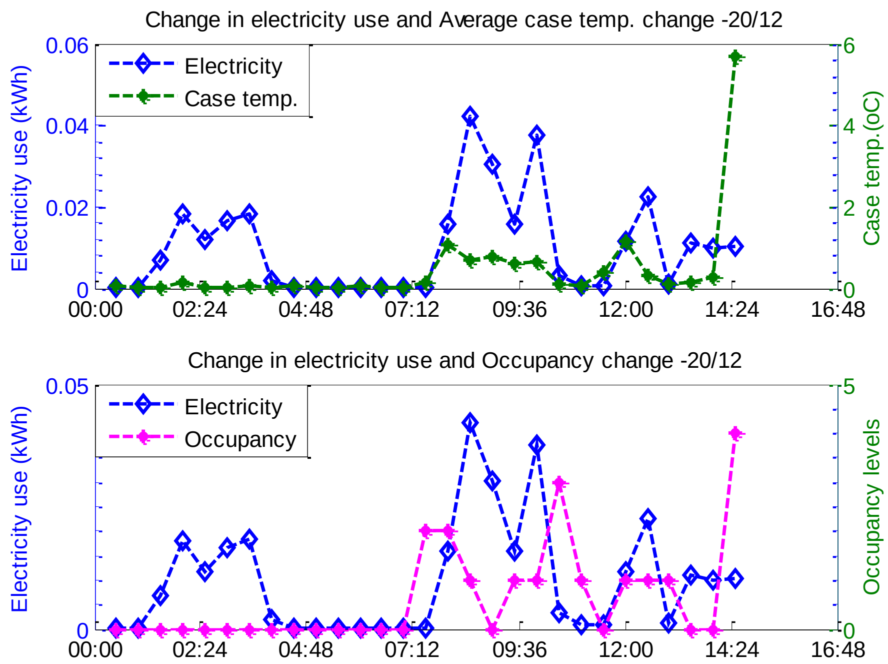


Figure (6.14): Change in electricity use, average case temperature, and occupancy levels-20/12

6.4 Occupancy-driven ventilation control and energy savings

Ventilation strategies using reliable real-time occupancy estimates can improve energy efficiency. Excess fresh air may have significant penalty on HVAC energy use depending on the outdoor temperature. Thus, keeping ventilation rates to the minimum required for acceptable IAQ based on real-time occupancy data has potential energy saving benefits. In order to test the practicality of the data–driven model developed in this research for energy savings, analysis was carried out for the determination of ventilation rates using two approaches: based on a fixed occupancy schedule, and occupancy estimates from the detection system. A detailed building ventilation design is beyond the scope of this thesis. A basic system was implemented for the purpose of illustrating energy savings with occupancy driven ventilation control.

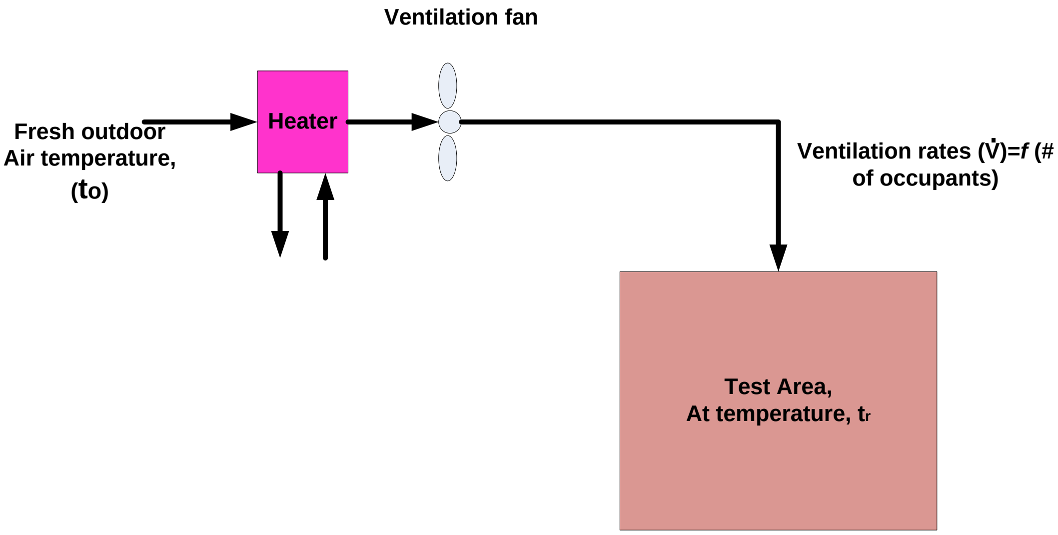


Figure (6.15): Basic ventilation system

A basic ventilation system comprising of a single fan (with a variable speed drive) and a heater, as shown in figure (6.15) was used to demonstrate energy savings for occupancy- driven building controls. The total energy (E) necessary for optimal ventilation is the energy consumed for heating the air mass (E_H) and that used for running the fan (E_V),therefore;

$$E = E_H + E_V \tag{6.1}$$

Ventilation requirements as per CIBSE guide B2 (CIBSE Guide B, 2001) were implemented in the analysis. Two basic requirements have to be satisfied for proper ventilation of a typical building space;

- Adequate supply of fresh air for the occupants
- Sufficient air changes in the space, so as to maintain satisfactory indoor air quality

System parameters

- **Fresh air ventilation rates**

Fresh air rate ($\dot{V}$) in m^3/s = Fresh air rate per person (l/s/p) x N (6.2)

Where N is the number of occupants in the space

The recommended sufficient ventilation rate per person is 8/l/s for office spaces (CIBSE, 2001). In this thesis, the maximum possible room occupancy at any given time was assumed to be 10 persons. Although the space accommodates 6 persons, it has a sitting capacity of 10. Hence, the minimum amount of fresh air required for satisfactory IAQ will be 80/l/s. Therefore, $\dot{V}$= 0.08 m^3/s or $288m^3/h$.

From figure (3.7), the space volume (V_L) = 8.003 x 13.006 x 3.5 = $364.30m^3$

Minimum supply air-change rate/ hour (ACH) = $\dot{V}$/ (V_L) = $288m^3/h$ / $364.30m^3$ = 0.79ACH. However, the recommended is 4.0ACH for a space accommodating up to 10 persons. Hence, the maximum airflow rate through the fan is (4 x 364.30) = $1457.20m^3/h$ or 404.7l/s.

- **Heating requirements**

During winter periods, introducing fresh outdoor air at low temperatures may be inconveniencing to occupants, hence the need to pre-heat the incoming air to about the internal room temperature. The energy required for heating an incoming supply of cold air is given as equation (6.3);

$$E_H = \rho_a \dot{V} C_{pa}(t_r - t_o) \tag{6.3}$$

Where ρ_a is the density of air, C_{pa} is the specific heat capacity of air, t_r is the indoor room temperature, and t_o is the outdoor temperature. ρ_a = 1.205kg/m^3 and C_{pa}= 1.005kJ/kg/K.

- **Fan power consumption**

The specific fan power (SFP_V) indicates the demand on the electrical power consumption of all supply air and extract air fans in a building. This determines the useful power applied for transporting air throughout an entire building. SFP_V of an individual fan in a ventilation system, such as that implemented in this research is given as in equation (6.4).

$$SFP_V = \frac{P_{MAX}}{\dot{V}_{design}} \tag{6.4}$$

Where P_{MAX} is the electrical power used by the fan (W) and $\dot{V}_{design}$ is the airflow rate through the fan (litre/s).

CIBSE guide B (CIBSE, 2001) suggests that using a SFP_V value of 2W/l/s is a good practice figure for ventilation fan specification in many buildings. This value was adopted in this thesis.

For a variable speed fan model, the flow characteristics for a fan with a variable speed drive are given by equation (6.5) and (6.6), as stated in energy plus (Energyplus, 2013b).

$$f_{flow} = \frac{\dot{V}}{\dot{V}_{design}} \tag{6.5}$$

$$f_{pl} = C_1 + C_2.f_{flow} + C_3.f^2{}_{flow} + C_4.f^3{}_{flow} + C_5.f^4{}_{flow} \tag{6.6}$$

Where C_1, C_2, C_3, C_4, and C_5 are the fan coefficient values at different fan speeds, f_{flow} is the flow fraction or part-load ratio and f_{pl} is the part-load factor. Table (6.1) presents the fan coefficient values.

The power consumed by the fan taking in to consideration possible pressure losses and non-linear fan characteristics is given by equation (6.7).

$$\text{Fan power consumption} = SFP_V \; x \; f_{pl} \tag{6.7}$$

Table (6.1): Fan coefficient values (Energyplus, 2013a)

Type of fan	Fan Coeff.1	Fan Coeff.2	Fan Coeff.3	Fan Coeff.4	Fan Coeff.5
Inlet vane dampers	0.35071223	0.30850535	-0.5413736	0.87198823	0.000
Discharge dampers	0.37073425	0.97250253	-0.3424076	0.000	0.000
Variable speed motor	0.0015302446	0.0052080574	1.1086242	-0.1163556	0.000

The ventilation rates for four representative days between 14/12/2012 and 19/12/2012 (excluding weekend days) are presented. Figures (6.16) – (6.19) show the ventilation rates profile using occupancy information based on fixed assumptions, model estimates and actual data. Ventilation rates were assumed to be constant during unoccupied times throughout the period examined. The potential for energy savings was investigated for occupied periods only. The space was maintained at an internal temperature of 21^{o}C during occupied times, and at 15^{o}C for unoccupied times.

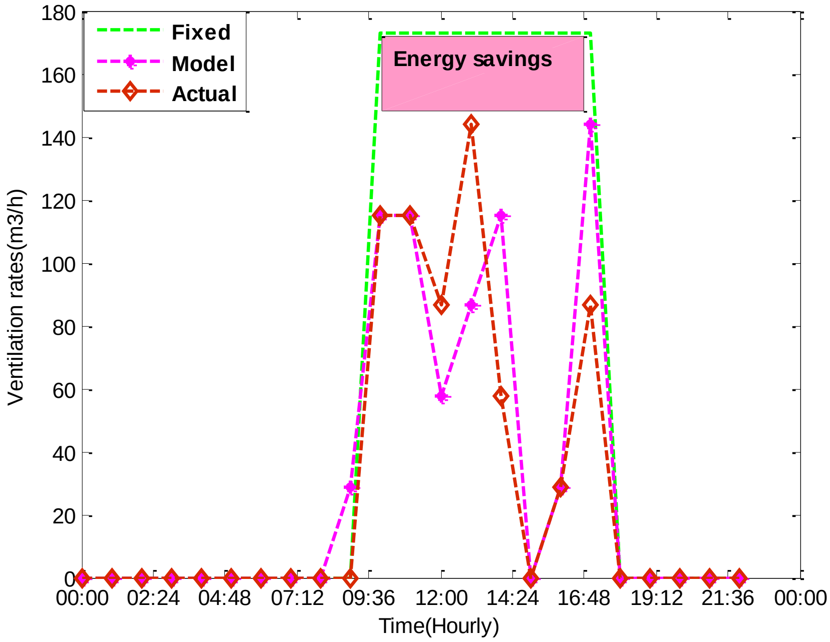


Figure (6.16): Ventilation rates using various control strategies for 14/12

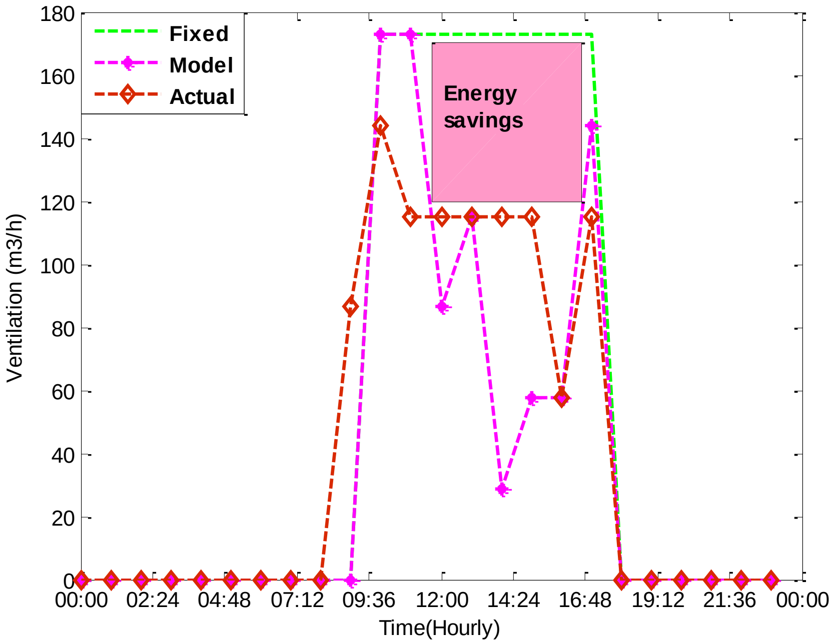


Figure (6.17): Ventilation rates using various control strategies for 17/12

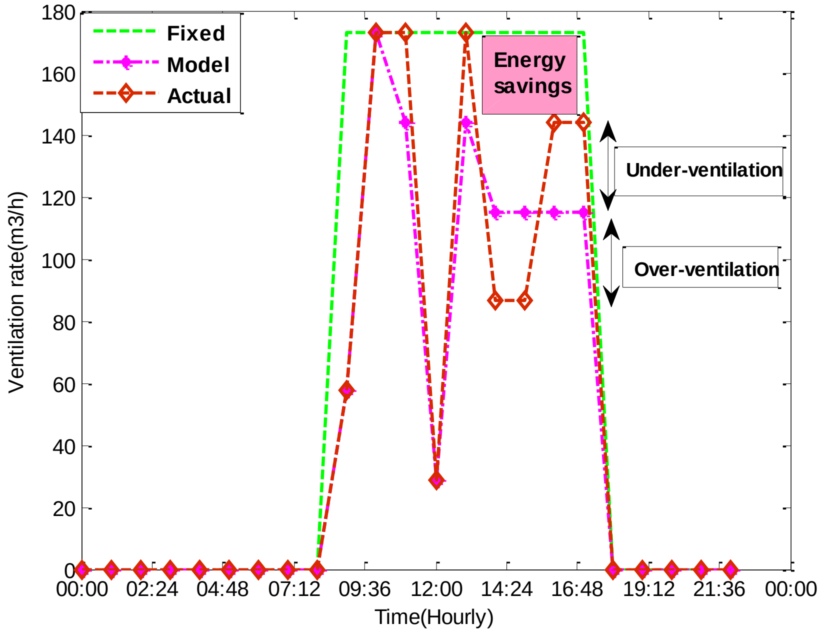


Figure (6.18): Ventilation rates using various control strategies for 18/12

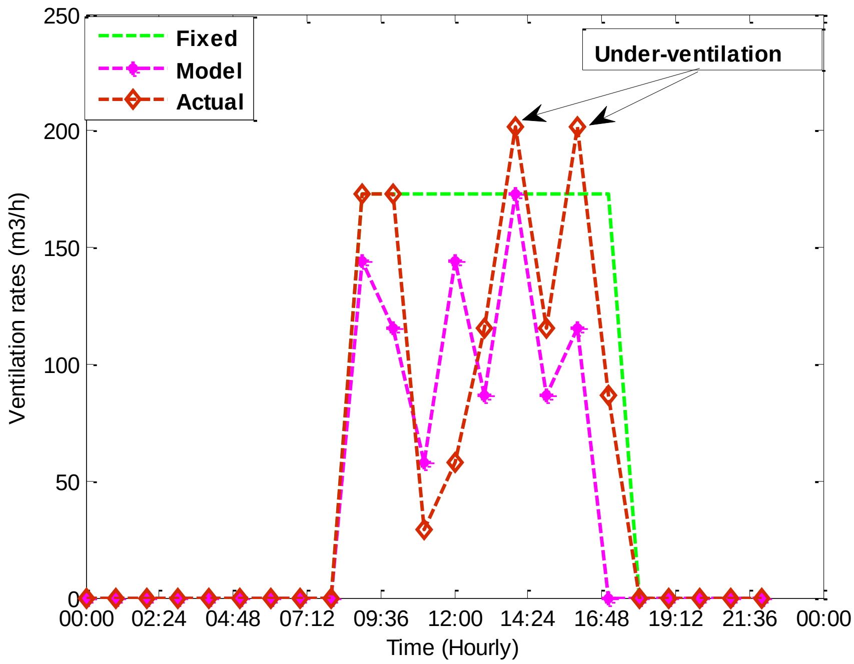


Figure (6.19): Ventilation rates using various control strategies for 19/12

The fixed schedule ventilation control strategy is based on maximum design occupancy expected during working hours, usually between 8am and 5:30pm. This schedule was kept constant during occupied periods, irrespective of occupancy levels. An occupancy profile of 6 was implemented in the fixed ventilation control analysis. In the other strategy, ventilation rates were a function of the model occupancy estimates and actual data.

Clearly from figures (6.16) – (6.19), the use of fixed schedules for ventilation control frequently overestimates the amount of ventilation required for maintaining satisfactory IAQ. However, with reliable occupancy information, the ventilation rates can be kept at optimal rates (just enough to provide an optimal flow rate of conditioned air into the space) or may be set back to minimum levels or turned off during unoccupied periods, resulting in possible energy savings whilst maintaining a comfortable environment.

Ventilation rates derived from estimated occupancy may sometimes indicate over ventilation or under-ventilation, causing energy waste or a decrease in comfort. This was influenced by the accuracy of the model estimation on each day. Deviations from actual occupancy levels may lead to excessive or insufficient fresh air supply, as illustrated in figure (6.18). Hence ventilation rates obtained using actual occupancy data were used to demonstrate energy savings, as they represent actual benefit of an occupancy driven control scheme.

Using fixed assumptions, the total volumetric fresh air required during occupied periods was 1382.40m^3/h corresponding to a total energy demand of 7.08kWh for heating the incoming air on the 14/12/2012, while using real-time occupancy data had a total fresh air and heating energy demand of 633.60m^3/h and 3.33kWh respectively, which amounted to a 53% energy savings for the day, see figure (6.20). The trend is similar for other days with potential energy savings of 27.28%, 31.34% and 28.44% shown for 17th, 18th and 19th/12/2012 respectively.

The ventilation fan electricity demand show a similar pattern, see figure (6.21), with possible energy savings of 53.77%, 28.96%, 31.26% and 28.16% for on the 14th, 17th, 18th and 19th respectively. Energy use obtained using model estimates were sometimes lesser compared to actual and fixed occupancy data, although this may be not always be optimal. Clearly, improved accuracy of occupancy detection system would benefit ventilation systems.

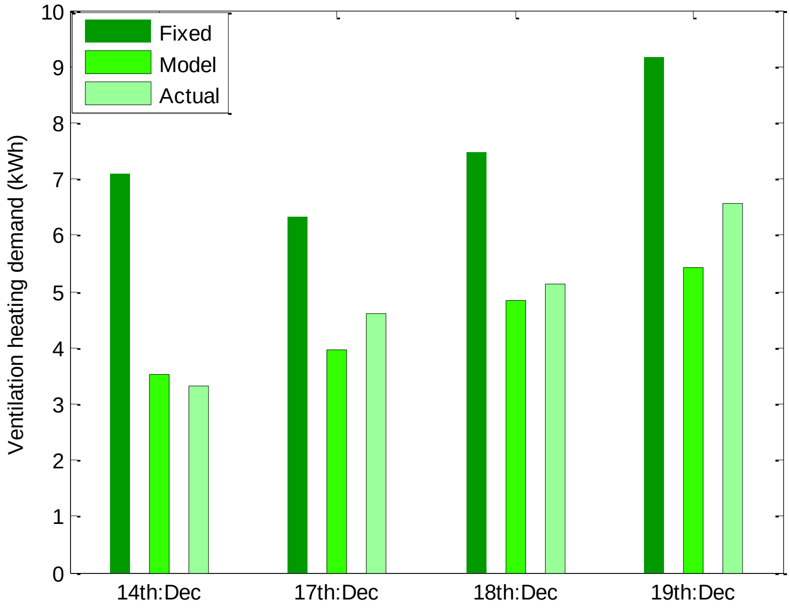


Figure (6.20): Heating demand using different ventilation rates

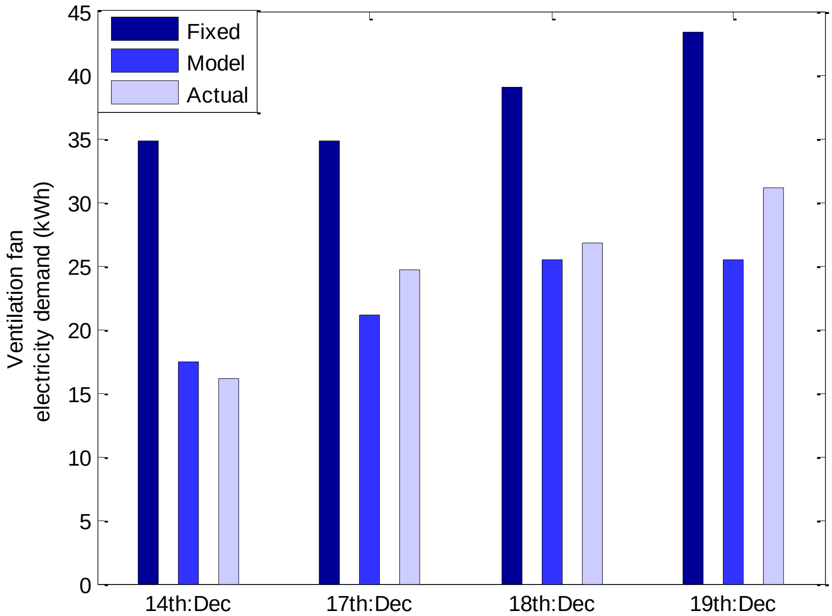


Figure (6.21): Ventilation fan electricity demand using different ventilation rates

6.5 Chapter summary

In this chapter, the relationship between various building variables such occupancy, indoor climate and energy (electricity) use was explored. VOC and CO_2 levels were strongly correlated in the space. Energy (electricity) use, VOC levels and average case temperature showed some relationship, although it was not clearly defined. The impact of real-time based occupancy-driven ventilation control strategy on building energy use was illustrated, with potential daily energy savings reaching 53% for both heating and electricity demand. Ventilation rates derived from actual and model occupancy data were smaller compared to those using fixed assumptions.

CHAPTER 7

DISCUSSIONS AND CONCLUSION

This chapter concludes the thesis by revisiting the research hypothesis and discussing key aspects of the research undertaken. The major contributions to knowledge, research significance and recommendations for future studies are presented. Finally, the research conclusions are drawn.

7.1 Research summary

In this study, the development and application of low-cost and non-intrusive multi-modal sensor networks for estimation of occupancy numbers has been presented. The system is shown to be capable of improving building energy management. In this section, the research summary and findings from the experiments carried out to test the research hypothesis are presented;

> *Research hypothesis: The combination of information derived from low-cost and non-intrusive indoor environmental sensors using machine learning techniques can provide reliable occupancy estimations in a naturally ventilated open-plan building.*

The successful implementation of any occupancy driven-HVAC strategy is largely dependent on the accuracy of the occupancy detection system deployed. However, current occupancy sensing technologies may limit the effectiveness of building controls, due to a number of issues ranging from unreliable data, sensor drift, privacy concerns, and the increased complexity of modern buildings, for example having integrated renewables, and tighter performance margins, compounds the difficulties imposed by pressure on commissioning. Besides, most occupancy detection systems rely on single monitoring point or use of a redundant sensor network or intrusive sensors to detect occupancy. From the reviewed literature, occupancy sensors developed from a machine learning based heterogeneous multisensory fusion strategy are seen to offer a more robust solution to challenges encountered in building occupancy monitoring. These sensors can facilitate improved system control performance, such that it is capable of turning off services out of hours, and not over-ventilating, thus enabling energy savings, and not under-ventilating during occupied periods, giving comfort and health benefits.

Despite these benefits, there is a shortage of methodology for developing robust and reliable occupancy monitoring systems. Very few studies in the literature have tackled the problem of occupancy numbers detection in non-domestic buildings, with as rich a sensing platform or numerous environmental ambient sensors, as deployed in this research. It is common to investigate the use of one or two type of sensors for occupancy monitoring, which are selected arbitrarily, and may not guarantee reliable occupancy monitoring. For instance, Dodier et al. (2006) utilised PIR sensors and a telephone sensor for occupancy detection, which can be prone to false switching of services. Table (2.1) provided a state-of-art summary and comparison of occupancy detection systems. These comparisons included type of sensors deployed, detection algorithms and features extracted. This research introduces a systematic methodology for selecting sensors and processing data for estimating occupancy levels. The proposed system aims to enhance the advantages of various sensors, whilst minimising their weaknesses by appropriately fusing information from the sensors. The proposed system can be adapted to work with an existing BEMS, such that sensor data can be fed in to the model, so as to determine an optimal combination of sensors capable of providing reliable occupancy monitoring.

The proposed occupancy detection strategy was carried out in two key stages; the hardware system implementation and data processing. The hardware system implementation included a custom made sound sensor and refinement of CO_2 sensor for EMI mitigation. Two test beds were designed and implemented for supporting the research studies, including proof-of-concept, and experimental studies. These test beds were deployed in two different locations within a naturally ventilated open-plan office building, with each gathering a range of indoor environmental climatic data. New sensing techniques were implemented in the multi-sensory instrumentation strategy implemented for occupancy levels estimation in chapter three. The custom low-cost sound sensor produces a binary output indicative of space occupancy and vacancy, without the use of any sophisticated analysis. The sound sensor was able to generate reliable and repeatable occupancy related information. The use of case temperature monitoring which has already been established as a useful method for detecting office equipment usage (Brown et al., 2011), was deployed to investigate its usefulness for occupancy levels monitoring. The EMI mitigation strategy involved shielding CO_2 sensors using a self-adhesive copper foil, ensuring there were no grounding loops while the power supply was provided by 6V 14Ah rechargeable lead-acid batteries to further mitigate mains bourne and power supply induced EMI.

In chapters four and five, novel data processing architecture for monitoring occupancy levels was introduced and tested. The data processing system integrated different stages such as features extraction, feature ranking, and feature selection. The feature extraction stage was intended to capture temporal parameter variations within the indoor environment. Features extracted from the sensor data included first and second order differences, variance and approximate area under the curve. The predictive capacities of different (features) sensors for occupancy levels estimation vary from one another. Therefore in order to achieve the best performance with an occupancy detection system, it was necessary to combine only sensors that provide maximum occupancy information. This action often requires the determination of the predictive ability of individual sensors, and an optimised selection process. For the feature ranking and selection stages, it was demonstrated in chapter four that the use of symmetrical uncertainty analysis and a genetic based search can identify an optimal sensor (features-subset) combination for occupancy numbers estimation. The most important sensors for occupancy numbers depend on the type of space under test, and the occupancy dynamics. Sound sensing was effective for both spaces tested in this research. Case temperature, sound and CO_2 sensors were important for a space with less air movement or rather with slow decay rates of indoor environmental parameters, while PIR and sound sensors were vital for an area with large volume and faster decay rates.

In chapter five, several test cases were experimented using various sensor network configurations to test and evaluate the effectiveness and feasibility of the proposed data processing methodology. A back-propagation neural network was adopted to combine candidate features for occupancy levels estimation in both test areas. The use of an optimal heterogeneous multisensory network out-performed any redundant sensor network for occupancy level monitoring, with estimation accuracy reaching 81.17% for occupied periods during week days. The complementary nature of a sensor fusion approach was evident in the analysis, such that the strengths of sensors make up for their weakness, thus ensuring enhanced performance. Sensor drop off can affect the performance of the occupancy detection system, with the exclusion of CO_2 sensor accounting for a 14.73% average reduction of the model accuracy as presented in section (5.7). In section (5.9), evaluation results of the proposed data processing methodology using other machine learning models confirm the potential and robustness of the proposed methodology for the development of an advanced building occupancy sensing system. However, a cross room estimation analysis (i.e. using the occupancy model trained from one room data to estimate occupancy in the other room) produced poor results (refer to section 5.10).

The relationship between VOC and CO_2 levels in the test area three was explained in chapter six. Findings from the analysis reveal both parameters have good correlation with R^2 value at 0.70. The implication of this is that both types of sensors may be used for similar functions in a ventilation control strategy. This can have profound impact on sensor selection and instrumentation cost for IAQ monitoring. However, more data from different building types and locations to carry out longitudinal analysis may be needed to establish this finding. Average case temperature measurements show some correlation with energy (electricity) use and VOC levels with R^2 value reaching up to 0.58 and 0.87 respectively. This helps to reinforce its usefulness for estimating occupancy levels in an office setting. The effectiveness of occupancy driven ventilation control was demonstrated by modulating ventilation rates in the space as a function of the real-time occupancy information, with the corresponding potential daily energy savings reaching 53% as shown in section 6.4.

In conclusion, the test results from the system are promising for building occupancy monitoring given that both test areas have a high ceiling, and are inside a naturally ventilated building. The proposed system addresses the issue of intrusion and privacy, since the sensors utilised in the study are non-intrusive and provide anonymous information. Besides, the system demonstrated is easy to deploy and can be adapted to form an overlay system which can be useful for BEMS commissioning.

7.2 Research Significance

Findings from this research can contribute toward efforts aimed at improving building operations, and can assist the UK government in achieving its target for climate change mitigation, by facilitating the lowering of building energy use. People spend about 80% of their life time inside buildings (Dounis and Caraiscos, 2009), and so the goal of any BEMS is to meet the comfort requirements of occupants in an energy efficient way. Roughly half of the total energy used up in buildings can be attributed to HVAC operations (Pérez-Lombard et al., 2008), so improving HVAC system operations can contribute to reductions in building energy use (Tachwali et al., 2007). Energy use in buildings is clearly linked with the activities of its occupants (Richardson et al., 2010), (Richardson et al., 2008), (Torriti, 2012). Thus, for monitoring and targeting to be successfully employed, so as to identify energy saving opportunities, occupancy detection becomes a crucial variable in energy management. More specifically, the impact of this research to some stakeholders in the built environment is discussed briefly under the following headings.

7.2.1 Demand side management (DSM)

Demand side management (DSM) refers to the actions undertaken at the demand side of energy meters; it basically explores ways to match demand with available supply within an energy (electricity) grid (Warren, 2014). DSM strategies may include time-of-use rates, real-time pricing, critical peak pricing and incentives (Warren, 2014). Energy use in buildings can be reduced by applying DSM solutions, by propelling changes in consumer’s behaviour, consequently matching demand with supply at lower energy prices (López-Rodríguez et al., 2013). Robust implementation of DSM solutions would require a comprehensive knowledge of energy (or electricity) use patterns, which can be clearly linked with active occupancy patterns, especially in residential buildings (Richardson et al., 2010),(Richardson et al., 2008), (Torriti, 2012). This may be so because during occupied periods, occupants are likely to indulge in activities that consume electricity such as lighting, utilising appliances etc. Occupancy data remain an important source of information for the determination of load profiles, especially in residential buildings (Abu-Sharkh et al., 2005).

In order to explore the potential for energy savings and reduced carbon emissions, it is crucial to identify demand peaks in electricity consumption patterns, and how these vary on a typical day (Torriti, 2012). Variations in peak occupancy can give insights to how flexible electrical loads might be during occupied periods (Torriti, 2012). In DSM strategies, where the main goal is to shift loads to off-peak hours, this operation should be based on the dynamics of peak occupancy profile (López-Rodríguez et al., 2013). Torriti (2012) also suggested that occupancy patterns (human activities in and out of the households) should be the sole factor informing DSM strategies in households, rather than electricity price, since it reflects the reality of electricity demand loads. Motuziene and Vilutiene (2013) recommended the use of maximum information of occupancy, and their preference for generation of demand profile rather than use fixed schedules, in order to ensure effective DSM strategies.

The proposed occupancy detection system can be utilised to generate active occupancy peaks, (days with high occupancy numbers), and the activities of the occupants in a DSM strategy. The corresponding time slots and occupancy variations of these peak days can be determined, and this information applied to develop advanced strategies to diminish peak demand, and thus reduce energy use. Case temperature monitoring utilised in this research can be useful to generate appliance usage profiles, which could be useful for power management strategies implemented in office buildings, as well as represent a starting point for generation of electricity demand profiles for the implementation of DSM strategies to smoothen peak loads, and thus facilitate reduction in energy use. Previous research has established that the operation of office equipment is not driven by indoor environmental comfort motives, hence that it was logical to link the ratio of internal heat gains over the nominal power of office equipment to the space occupancy dynamics (Parys et al., 2011).

In large scale non-domestic buildings where occupancy patterns may change frequently, the use of a software routine to ping office equipment connected to the local network in the space as implemented in Brown et al. (2011), may produce similar results as that of case temperature monitoring. These methods hold potential for reducing the cost of building instrumentation in gathering information about space and electronic appliance usage, which can be beneficial to HVAC and DSM operations.

Clearly, reliable knowledge of real-time occupancy is beneficial to DSM solutions. Hence, reliability and robustness of the occupancy detection system deployed in any DSM strategy becomes crucial.

7.2.2 Architects/ Designers

During the design phase of buildings, simulation tools (such as EnergyPlus, IES, eQuest etc) are usually utilized to predict energy use, and also help designers in equipment selection (Hoes et al., 2009). In reality, energy predicted using these tools often differ from the actual energy use during the operational life of the building, with typical average variation reaching 30% (Yudelson, 2010), (Turner and Frankel, 2008). Such deviations may be attributed in part to occupancy, although building simulation tools rarely consider the impact of occupants' behaviour on energy use (Azar and Menassa, 2012). Several studies have emphasized the need to properly account for occupancy parameters in simulations, so as to improve the sensitivity of a building model, and to generate more reliable building energy performance predictions (Peschiera et al., 2010), (Azar and Menassa, 2012). Occupancy information derived using this methodology can be useful for validation of building simulation models, simulating new buildings control strategies based on realistic occupancy levels, as opposed to using fixed assumptions.

Achieving the full potential of occupancy-driven HVAC operations may require more than simply running system controls based on indication of space occupancy or vacancy. Typical HVAC and building equipment are often designed based on steady state operations, including occupancy considerations. These systems struggle to adapt to the building occupancy dynamics, in terms of energy efficiency and comfort requirements. There is clearly a need for the design of HVAC equipment than can cater for dynamic occupancy schedules in different building configurations. Understanding occupancy patterns in buildings would be a prerequisite in the design and development of any energy efficient HVAC system (Whitehouse et al., 2012). The proposed occupancy system contributes to efforts aimed at generating dynamic and reliable building occupancy schedules.

7.2.3 Building controls industry

There is a disconnect between the market and scientific research in this area, such that installed building controls in most commercial buildings are still miles behind other advanced controllers (Dounis and Caraiscos, 2009). Existing building controls rarely make use of an occupancy variable as an input for system control beyond an abstracted figure from atmospheric gases or a simple binary value. Whereas, advanced controllers which are capable of this integration have not been commercialized (Dounis and Caraiscos, 2009). Again as mentioned earlier in section (7.2.1), occupants' behaviours and activities have significant impact on energy use (Richardson et al., 2010), (Richardson et al., 2008); it therefore becomes important for reliable occupancy information to be an input for system control. The industry will need to move fast to close this gap in building energy efficiency.

In the future, BEMS may be expected to monitor hourly electricity market prices such as to avoid peak loads (Yang and Wang, 2013). Hence, it may be useful for existing BEMS to be more responsive, in order to provide better pricing information to end-users, alongside maintaining optimal control in the environment. For buildings to be more adaptable to occupants comfort needs, whilst maximizing energy savings, the concept of ambient intelligence should be embedded in to the controlled space such that the environment is aware of the users' activities (Nguyen and Aiello, 2013). Virtual occupancy sensing such as the one discussed throughout this thesis can be further developed to recognise occupant's activities, and adapt to changes within an indoor environment to ensure energy efficient building operations. It can also one day deliver a seamless integration of building controls and occupancy information. However, the development of a hardware and software interface will be crucial for this task. The conceptual study of ventilation control based on occupancy data shown in this thesis (section 6.4), generated potential energy savings of up to 53% in a day. However, the next steps will be to confirm its benefits in real-life ventilation installation.

7.2.4 Building services

In building energy management, low-cost and non-intrusive sensor networks can be vital for continuous and seamless monitoring of occupancy profiles, with a view to reduce energy use. These networks can provide a basic platform for gathering information on occupants' behaviours, as well as their interactions with their indoor environments. For instance, the proposed occupancy detection system can be utilised to generate different occupancy, appliance usage and behavioural profiles for different spaces in an open-plan office building. Although the proposed occupancy sensing network makes use of information from CO_2, sound, PIR and case temperature sensors (as per two test areas examined), the system can be flexible to accommodate other sensing mechanisms, to examine their feasibility for occupancy monitoring.

Active occupancy patterns generated from the proposed system can be applied to maintain optimal control set –points and operating modes of HVAC systems to facilitate energy savings, as implemented in various demand-driven HVAC and lighting strategies in the chapter two (Li et al., 2012), (Tiller et al., 2010). Besides this, the system also monitors other indoor environmental variables simultaneously (such as air quality and internal temperature).

Huge financial resources are wasted managing building operations due to faulty building controls (Brown et al., 2010). The functionality of many sensors can be affected by the environment in which they have been deployed. For instance, as described in chapter three (section 3.3) of the thesis, the presence of RF sources in the test building introduced EMI to the measurement circuits of CO_2 sensors, whose outputs were plagued with noisy harmonics. If such a sensor is deployed as a control sensor for ventilation purposes, this could have an adverse impact on energy efficiency and air quality. As demonstrated in chapter three (section 3.3) of the thesis, good analogue designs for BEMS sensors can address such a problem. Continuous commissioning of BEMS sensors with the use of virtual sensors may offer potential benefits in fault diagnosis (Li et al., 2011), which may go a long way in improving the performance of BEMS.

The short-comings of VOC and CO_2 sensing for ventilation control were highlighted in chapter two (section 2.4.6). For instance, CO_2 sensors have a long response time, such that by the time sensors detect high levels of CO_2 that trigger ventilation, occupants may already be in a state of discomfort (Fisk, 2008). It becomes useful to explore the feasibility of alternative low-cost sensors (such as the sound sensor developed in this research) to assist in a demand-driven ventilation strategy. This sound sensor network can be utilised to compliment information input to a ventilation control system, as a result of the slow response time for CO_2 sensors to detect incoming occupants.

In many instances where there is change of use for a particular building space, sensors may not be changed and this may have ramifications for climate control. Resilience in building control sensor networks becomes an important issue with respect to the performance of any installed BEMS. For instance, if an office space is turned in to a kitchen, naturally VOC levels will be expected to rise. Where control sensors are not replaced so as to cope with the change in ventilation loads, or where the BEMS does not possess a self-tuning capacity, the indoor air quality may likely be compromised. This may serve as a motivation for integrating a resilient sensor fusion network with BEMS. Self –tuning model for BEMS have been applied for online tuning HVAC parameters to facilitate optimal control and energy savings (Nassif et al., 2008). In section (5.7) of the thesis, sensor resilience was examined by simulating sensor's drop-off, with that of CO_2 accounting for 14.73% average reduction in the accuracy of occupancy estimation.

With widespread practice of deploying dense spatial instrumentation points, any methodology to optimise sensor selection and placement for comfortable and energy efficient environment would be very useful. Replacing large number of sensors with a couple of more powerful low-cost and non-invasive sensors, such as those developed from a sensor fusion process can have compelling implications for the cost of building instrumentation. However, this could be demonstrated in any future work.

7.2.5 Society

Other wider benefits of this research to the society are discussed briefly under the two following headings; Care management for the elderly and public buildings.

- **Public buildings**

People spend significant amount of time in public buildings (such as shopping mall, airports, train stations etc), and usually moves between various service facilities. Occupancy movement in such locations need to be carefully analysed and reliably predicted in case of emergency situations, and generally for the provision of an optimal level of service (Nassar, 2010). Occupants flow information can be useful to building managers in diverting occupants traffic, allocation of rooms, facilitates and service zones (Lee et al., 2012). Occupants flow in public buildings have been determined using mathematical models (Tabak and de Vries, 2010), although they may require significant time and effort to develop. Video sensing is also widely used (Li et al., 2009a), but this can be cost intensive. An interesting alternative can be to monitor occupants flow in these spaces from indoor environmental variables using low-cost and non-invasive sensors. The proposed occupancy detection system can be adapted to provide occupancy counts/unit space area, and this information can be applied to construct occupants flow models for different locations within the building. This can be then utilised for optimisation of space use, and also ensuring that building services are kept at satisfactory conditions.

Besides monitoring of occupancy flow density or traffic, maintaining acceptable indoor air quality (IAQ) is seen to be important for obvious reasons. However, the IAQ in public buildings such as shopping malls can quickly deteriorate during peak times (Li et al., 2001). Thus, a system capable of detecting occupancy levels as well as IAQ becomes extremely useful for any occupancy-driven ventilation strategy.

- **Care management for the elderly**

Another potential application of the proposed occupancy detection system is that of elderly care management. Context activity –awareness is seen as important for care management of the elderly, such that deployment of ubiquitous sensor networks becomes necessary (Alemdar and Ersoy, 2010). Wearable sensors are commonly used in this area (Korhonen et al., 2003), (Lötjönen et al., 2003), although the attendant issue of intrusion and acceptability remains a concern. The proposed occupancy sensing platform can be adapted and deployed for providing contextual information obtained from fusion of various environmental sensors. The network may be applied to track occupants' movement, and monitor their daily activities such as appliance usage. These sort of information are useful for elderly care management (Alemdar and Ersoy, 2010).

7.3 Contribution to knowledge

The main contributions made by this thesis are outlined as follows;

- Occupancy detection in buildings has remained difficult, partly due to the lack of a well-defined process for sensor selection to reliably gather occupancy information, as well as short comings of existing occupancy detection technologies. These issues have been well elucidated in chapter two (sections 2.1.1, 2.1.2, and 2.2.1) of this thesis. Arbitrary selection of sensors for occupancy estimation does not guarantee robust performance. This research introduced an innovative and systematic data processing system (refer to figure 4.1), to facilitate optimal (relevant) sensor selection for the development of a robust and reliable occupancy estimation system.

  For the data collected in this research, where unoccupied instances produce "zero value" for significant periods (for motion and acoustic measurements), the use of state-of-the-art technique such as information gain can be biased toward such features with more values when applied for determination of the predictive strength of individual sensor features. Symmetrical uncertainty analysis, which does not have the problem of bias, was used to determine the predictive capacity of various features (sensors). In order to arrive at an optimal feature subset for fusion in the occupancy estimation model, a genetic search algorithm was employed. This is computationally less demanding compared to a technique such as exhaustive search applied in occupancy detection systems in the literature. A neural network was implemented for combination of candidate features for occupancy numbers detection.

The systematic combination of these advanced processing techniques is completely new for the development of occupancy detection systems. This innovative system of processing indoor environmental data has the potential for utilising different sensors inputs from an existing BEMS within a specific environment for reliable occupancy monitoring, such that only the relevant ones are utilised, and therefore can impact on the instrumentation requirements for building performance monitoring. This sort of flexibility forms a unique strength of the proposed system, as most occupancy detection system in the literature lacks this quality. Extensive tests provide evidence that this novel methodology is able to provide reliable estimation results with accuracy reaching 84.59% during occupied instances, and also capable of supporting occupancy-driven ventilation control strategy, although this can be extended to heating and cooling systems.

- New techniques for building instrumentation were introduced to extract occupancy related information from the test areas. The use of case temperature monitoring was identified as a useful method for monitoring occupancy numbers in open-plan office buildings, although this may be subject to specific conditions, such as all the desktop computers in the space being instrumented. A new low-cost sound sensor which does not require any sophisticated analysis to extract information for occupancy numbers estimation was designed and applied in this study. This sound sensor works like a typical PIR sensor, producing binary outputs for vacancy and occupancy intervals. Convectional sound sensors such as boundary microphones commonly deployed for acoustics monitoring in buildings are relatively more expensive (with the cheapest ones selling at around £35) compared to the one applied in this research. The sound sensor in this research cost about £10 to assemble, and this price could be further reduced if mass produced. Test accuracy for the sound sensor confirms its use as a promising tool for monitoring occupancy levels.

In general, for development of robust occupancy detection systems, it is quite crucial that the deployed sensors capture occupancy information in a non-invasive and anonymous manner, such that occupants' activities are not disrupted. Both instrumentations are low cost, non-intrusive and completely new in terms of application for occupancy numbers detection.

- Few studies in the literature have investigated the issue of occupancy detection in office buildings with a similar network of sensors as the type implemented in this study. To the best of the researcher's knowledge, none of these studies were carried out in a naturally ventilated space, where the indoor climate can be quite dynamic due to the complex air flow regimes in the building, which could impact the performance of an occupancy detection system. A significant aspect of this work is that the environmental data used for occupancy levels estimation were collected from a naturally ventilated building. This study demonstrated that the methodology adopted works efficiently in this type of setting.

7.4 Recommended future work

The following issues have been identified for further research efforts, but being beyond the scope of this study;

- As mentioned earlier in chapter two (section 2.4), BEMS rarely make use of occupancy information as a system input for control of building services other than in lighting control. Further work to develop a commercial platform for integrating real-time occupancy data into HVAC system control would be a welcome development to the industry, as research efforts in this area are still in the early stages.

- For occupancy driven control, apart from the development of a reliable system for occupancy level monitoring, a crucial aspect is the determination of occupancy patterns (duration and the frequency of the time interval of space occupancy). The development of a feedback system to compensate for performance variation in an occupancy detection model, due to the influence of outdoor climatic conditions would be useful. The above issues are relevant to the performance evaluation of any building control strategy, and would require further investigation.
- Since the results achieved are limited to the specific environment observed in this study, extension of the applicability of this research to other building configuration different from a typical naturally-ventilated open-plan office setting would prove useful in establishing its robustness. It would be interesting to test the appropriateness and robustness of the proposed data processing methodology for occupancy monitoring in public places such as shopping malls, schools, hospitals, subway stations, airports, sports and gymnasium centres.
- Further work would be necessary to evaluate optimal location of sensors, standardising the number of sensors deployed per unit area and the associated potential savings in instrumentation cost (in unit such as kWh/m^2/area) for reliable occupancy sensing. Such information can assist building or facilities managers to make informed decisions about occupancy sensor selection early on during procurement, such that instrumentation cost can be brought down, whilst maintaining acceptable indoor environment for a particular building type.
- Unlike physical sensors whose performances are evaluated by accuracy, repeatability, sensitivity, etc, there are currently no standard methods for evaluation of a virtual sensor performance over a wide range of operational conditions. Most virtual sensors are often based on outputs from physical processes which can change over time due to the development of anomalies

or faults. It may be necessary to consider the robustness of virtual sensor outputs with respect to the development of physical faults in any performance evaluation methodology. Hence, there is clear motivation to include this in any further research.

7.5 Conclusion

With roughly about half of the energy used in buildings attributed to HVAC systems, there is clearly great potential for energy saving through improved building operations. Accurate knowledge of localised and real-time occupancy numbers can have compelling control applications for HVAC systems. However, existing technologies applied for building occupancy measurements are limited, such that a precise and reliable occupant count is difficult to obtain. An extensive literature review on state-of –the-art in building indoor environmental monitoring and control (focussing on occupancy detection) clearly demonstrated the need for a heterogeneous multisensory fusion approach in the development of occupancy detection systems. This approach seems to have the best potential in addressing the short-comings of existing occupancy detection systems. For system development, an experimental design utilising assorted low-cost indoor environmental sensors deployed in open-plan office environment was applied to demonstrate proof-of-concept. Data processing techniques and system hardware in the proposed occupancy detection system were developed for this purpose. The instrumentation strategy employed in the experimental design included the design of a custom sound sensor and an EMI mitigation strategy for CO_2 sensors. Experimental results confirmed that indoor environmental measurements are capable of providing repeatable occupancy related information.

This thesis has proposed an advanced data processing methodology to detect occupancy numbers in naturally ventilated open-plan office buildings based on information from a network of low-cost sensors. Various stages in the methodology included pre-processing, features extraction, SU based feature ranking, and correlation-based feature selection. This involved the use of symmetrical uncertainty analysis and a genetic based search for feature selection, and a machine learning model for sensor data fusion. The proposed approach utilised information from CO_2, PIR, sound and temperature sensors attached to office equipment, which was then combined with a neural network for estimating occupancy levels. Extensive experimental testing demonstrated that the system offers a promising approach for building occupancy sensing capable of supporting improved HVAC operations, and thereby facilitating energy savings. In general, the model estimations tracked actual occupancy levels with accuracy reaching 84.59% during occupied instances, which compares favourably with existing occupancy detection systems performance. However, the model works best for the data trained on. The impact of real-time based occupancy-driven ventilation control strategy on building energy use was illustrated, with potential daily energy savings reaching 53% for both heating and electricity demand. It is therefore reasonable to conclude that the aim and objectives set forth at the start of the study have been achieved.

REFERENCES

ABU-SHARKH, S. R., MARKVART, T., ROSS, N., WILSON, P., YAO, R., STEEMERS, K., KOHLER, J. & ARNOLD, R. 2005. *Microgrids: Distributed on-site generation, TechnicalReport 22. TyndallCentreforClimateChangeResearch.* [Online]. Available: http://www.tyndall.ac.uk/sites/default/files/it1_33.pdf [Accessed 03/08/2013].

ABUSHAKRA, B. & CLARIDGE, D. E. 2008. Modeling Office Building Occupancy in Hourly Data-Driven and Detailed Energy Simulation Programs. *ASHRAE Transactions*, 114, pp. 472-481.

AGARWAL, Y., BALAJI, B., DUTTA, S., GUPTA, R. K. & WENG, T. 2011. Duty-cycling buildings aggressively: The next frontier in HVAC control. *In: Proceedings of the 10th International Conference on Information Processing in Sensor Networks (IPSN)*, 12-14 April, pp. 246-257.

AGARWAL, Y., BALAJI, B., GUPTA, R., LYLES, J., WEI, M. & WENG, T. 2010. Occupancy-Driven Energy Management for Smart Building Automation. *In: Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building*, 03-05 November, Zurich, Switzerland.

AGLAN, H. A. 2003. Predictive model for CO2 generation and decay in building envelopes. *Journal of Applied Physics*, 93, pp. 1287-1290.

AGNELLO, S. 1999. Limitations of VOC sensors in Achieving Adequate Indoor Ventilation. *Environmental Newsletter.* Chelsea Group Limited, HI, USA.

AL-HABAIBEH, A. S., F; BROWN, N; KERR, D; JACKSON, M. 2004. A Novel Approach for Quality Control System Using Sensor Fusion of Infrared and Visual Image Processing for Laser Sealing of Food Containers. *Measurement,Science and Technology*, 15, pp. 1995-2000.

ALCALÁ, R., CASILLAS, J., CORDÓN, O., GONZÁLEZ, A. & HERRERA, F. 2005. A genetic rule weighting and selection process for fuzzy control of heating, ventilating and air conditioning systems. *Engineering Applications of Artificial Intelligence*, 18, pp. 279-296.

ALEMDAR, H. & ERSOY, C. 2010. Wireless sensor networks for healthcare: A survey. *Computer Networks*, 54, pp. 2688-2710.

ANDERSSON, K., BAKKE, J. V., BJØRSETH, O., BORNEHAG, C. G., CLAUSEN, G., HONGSLO, J. K., KJELLMAN, M., KJÆRGAARD, S., LEVY, F., MØLHAVE, L., SKERFVING, S. & SUNDELL, J. 1997. TVOC and health in non-industrial indoor environments. *Indoor Air*, 7, pp. 78-91.

ANDERSSON, M. & ILESTRAND, M. 2007. Data fusion of secondary and primary surveillance radars for increased robustness in air-traffic monitoring. *In: Proceedings of Radar Conference*, 9-12 October, Munich, pp. 456-459.

ANON. 2005a. *Elster valve mounted meter* [Online]. Available: http://www.elstermetering.com/en/853.shtml [Accessed 03/12/2010].

ANON. 2005b. *Thermocouples vs. Thermistors - Which are best for Thermal Validation?* [Online]. Veriteq Instruments Inc. Available: http://www.veriteq.com/validation/thermocouples-vs-thermistors.htm [Accessed 09/09/ 2011].

ANON. 2005c. *Universal Light Monitor & Data Logger* [Online]. Available: http://www.elsec.co.uk/ [Accessed 05/04/2011].

APTE, M. G. 2006. A review of demand controlled ventilation. *In: Proceedings of Healthy Buildings* , 4-8 June, Lisboa, Portugal, pp. 371-376.

ARMSTRONG, J. & COLLOPY, F. 1992. Error Measures For Generalizing About Forecasting Methods: Empirical Comparisons. *International Journal of Forecasting*, 8, pp. 69-80.

ASHRAE 2004. ASHRAE Standard 62.1: Ventilation for Acceptable Indoor Air Quality.

ASHRAE 2007. ASHRAE 90.1 Standard: Energy Standard for Buildings Except Low-Rise Residential Buildings.

AVOR, J. K. & SARKODIE-GYAN, T. 2009. An approach to sensor fusion in medical robots. *In: Proceedings of Rehabilitation Robotics, ICORR 2009, IEEE International Conference,* 23-26 June, Kyoto, Japan, pp. 818-822.

AXIOMATIC-TECHNOLOGY-LIMITED. *Beam counters* [Online]. Available: http://www.peoplecounting.co.uk/about-us.html [Accessed 30/01/2013].

AZAR, E. & MENASSA, C. C. 2012. A comprehensive analysis of the impact of occupancy parameters in energy simulation of office buildings. *Energy and Buildings*, 55, pp. 841-853.

AZZAM, H., KNIGHT, P., COOK, J. & WAKEFIELD, N. 2005. FUMS™ fusion for improved aircraft MAAAP. *In: Proceedings of IEEE Aerospace Conference*, 5-12 March, Big Sky, MT, pp. 3782-3796.

BEDWORTH, M. 1994. Probability Moderation for Multilevel Information Processing. *DRA Technical Report, DRA/CIS(SE1)/651/8/M94.AS03BP032/1*.

BEDWORTH, M. & O'BRIEN, J. 1999. The Omnibus Model: A New Model of Data Fusion? *Aerospace and Electronic Systems Magazine, IEEE* , 15, pp.30-36.

BENEZETH, Y., LAURENT, H., EMILE, B. & ROSENBERGER, C. 2011. Towards a sensor for detecting human presence and characterizing activity. *Energy and Buildings*, 43, 305-314.

BERRIOS, I. T., ZHANG, J. S., GUO, B., SMITH, J. & ZHANG, Z. 2003. *Volatile Organic Compounds (VOCS) Emissions from Sources in a Partitioned Office Environment and Their Impact on IAQ* [Online]. Available: http://beesl.syr.edu/pdf/Officeenvironment-abstract.pdf [Accessed 04/05/2012].

BERTRAND, P. 2001. *Method and apparatus for measuring the consumption of an element of an electrical network*. France patent application. [Online]. Available: http://patent.ipexl.com/EP/1136829ZZDASHZZA1.html [Accessed 10/05/2011].

BILGEHAN, M. 2011. Comparison of ANFIS and NN models—With a study in critical buckling load estimation. *Applied Soft Computing*, 11, pp. 3779-3791.

BISHOP, R. H. 2002. *The Mechatronics Handbook*, CRC Press.

BLUM, A. & LANGLEY, P. 1997. Selection of relevant features and examples in machine learning. *Artificial Intelligence*, 97, pp. 245-271.

BOAIT, P. J. & RYLATT, R. M. 2010. A method for fully automatic operation of domestic heating. *Energy and Buildings*, pp. 42, 11-16.

BOYD, J. 1987. *RE: A Discourse on Winning and Losing.*

BROWN, N. 2010. Image Processing for Overnight Lighting Quantification in Buildings. *In: Proceedings of Improving Energy Efficiency in Commercial Building Conference (IEECB 2010),* 13-14 April, Frankfurt, Germany.

BROWN, N., BULL, R., FARUK, F. & EKWEVUGBE, T. 2011. Novel instrumentation for monitoring after-hours electricity consumption of electrical equipment, and some potential savings from a switch-off campaign. *Energy and Buildings*, 47, pp.74-83.

BROWN, N. & WRIGHT, A. J. 2008. *Non Invasive and Cost Effective Monitoring of Energy Consumption Patterns for Electrical Equipment* [Online]. Available: http://www.ucl.ac.uk/carb/pubdocs/CP-DMU-09-IEECB08-NoninvasiveMonitoring-2008-NRB-AJW.pdf [Accessed 12/04/2011].

BROWN, N., WRIGHT, A. J., SHUKLA, A. & G. STUART 2010. Longitudinal analysis of energy metering data from non-domestic buildings. *Building Research & Information*, 38, pp. 80-91.

BSI 1999. EN 62053-61:1998 Electricity metering equipment (a.c.) —Particular requirements. *EN 62053-61:1998.* BSI.

BSI 2002. Electricity metering. Glossary of terms, PD IEC TR 62051:1999, IEC TR 62051:1999. BSI.

BURNHAM, K. P. & ANDERSON, D. R. 2002. *Model selection and multimodel inference : a practical information-theoretic approach,* New York.

BYRD, R. & NOCEDAL, J. 1989. A tool for the analysis of quasi-Newton methods with application to unconstrained minimization. *SIAM Journal on Numerical Analysis*, 26, pp. 727-739.

BYRD, R., NOCEDAL, J. & YUAN, Y. 1987. Global convergence of a class of quasi-Newton methods on convex problems. *SIAM Journal on Numerical Analysis*, 24, pp. 1171-1189.

C.E.C 2008. Strategic Plan to Reduce the Energy Impact of Air Conditioners. Report CEC 400-2008-010. California Energy Commission.

CALOGIROU, A., BOEKHOVEN, J. & HENKES, R. 2001. Effect of wall roughness changes on ultrasonic gas flowmeters. *Flow Measurement and Instrumentation*, 12, pp. 219-229.

CALVINO, F., LA GENNUSA, M., RIZZO, G. & SCACCIANOCE, G. 2004. The control of indoor thermal comfort conditions: introducing a fuzzy adaptive controller. *Energy and Buildings*, 36, pp. 97-102.

CASCETTA, F. & VIGO, P. 1994. The future domestic gas meter: Review of current developments. *Measurement*, 13, pp. 129-145.

CHEN, B. & CHEN, S. 2010. Multisensor Information Fusion in Pulsed GTAW based on Fuzzy Measure and Fuzzy Integral. *Assembly Automation*, 30**,** 276-285.

CHEN, T., CHEN, T. & CHEN, Z. 2006. An Intelligent People-Flow Counting Method for Passing Through a Gate. *In Proceedings of IEEE Conference on*

Robotics, Automation and Mechatronics, 1-3 June, Bangkok, Thailand, pp. 1-6.

CHEN, T. Y. 2001. Real-time predictive supervisory operation of building thermal systems with thermal mass. *Energy and Buildings*, 33, pp.141-150.

CHENDA, L. & BAROOAH, P. 2010. An integrated approach to occupancy modeling and estimation in commercial buildings. *In: Proceedings of American Control Conference (ACC), 2010*, 30 June - 2 July, Baltimore, USA, pp. 3130-3135.

CHEONG, K. W. 2001. Airflow measurements for balancing of air distribution system — tracer-gas technique as an alternative? *Building and Environment*, 36, pp. 955-964.

CHERKASSKY, Y. & MA, Y. 2004. Practical selection of SVM parameters and noise estimation for SVM regression. *Neural Networks*, 17, pp. 113-126.

CHOU, J. 2000. *Hazardous gas monitors: A pratical guide to selection, operation and applications*, McGraw-Hill Book Company.

CHRISTENSEN, K. J., GUNARATNE, C., NORDMAN, B. & GEORGE, A. D. 2004. The next frontier for communications networks: power management. *Computer Communications*, 27, pp.1758-1770.

CHU, C., JONG, T. & HUANG, Y. 2005. Thermal comfort control on multi-room fan coil unit system using LEE-based fuzzy logic. *Energy Conversion and Management*, 46, pp. 1579-1593.

CIBSE .1996. CIBSE Commissioning Code A: Air Distribution Systems. Chartered Institiue of Building Services Engineers, London , United Kingdom

CIBSE .2001. Heating, Ventilating, Air Conditioning and Refrigeration (CIBSE Guide B). Chartered Instititue of Building Services Engineers, London , United Kingdom.

CIBSE .2009. Building Control System. Chartered Instititue of Building Services Engineers, London , United Kingdom.

CINCON. *TRG500 SERIES 6W Switching Adapter* [Online]. Available: http://www.cincon.com/data/products/adcd1_3/TRG500.pdf [Accessed 20/03/2012].

CISCO. 2005. *Cisco connected real estate.* [Online]. Available: http://www.builconn.com/downloads/CCRE-WP2.pdf [Accessed 10/10/2013].

CLEVELAND, M. A. & SCHUH, J. M. 2010. Automating the residential thermostat based on house occupancy. *In: Proceedings of IEEE Systems and Information Engineering Design Symposium (SIEDS)*, 23-23 April, Charlottesville, VA, USA, pp. 36-41.

COHEN, D., KRARTI, M. 1995. A Neural network approach applied to energy conservation retrofits *In: Proceedings of 1995 Building Simulation Conference*, August, Madison, Wisconsin , USA.

COHEN, R., STANDEVEN, M., BORDASS, B. & LEAMAN, A. 2001. Assessing building performance in use 1: the Probe process. *Building Research and Information*, 29, pp. 85-102.

COVER, T. M. & THOMAS, J. A. 1991. *Elements of Information Theory*, Wiley.

CULLEN, J. D., ATHI, N., AL-JADER, M., JOHNSON, P., AL-SHAMMA'A, A. I., SHAW, A. & EL-RASHEED, A. M. A. 2008. Multisensor fusion for on line monitoring of the quality of spot welding in automotive industry. *Measurement*, 41, pp. 412-423.

CURTIS, P. S., SHAVIT, G., KREIDER, K., 1996. Neural networks applied to buildings-a tutorial and case studies in prediction and adaptive control *ASHRAE Transactions* 102, pp.1141-1146.

DALAMAGKIDIS, K., KOLOKOTSA, D., KALAITZAKIS, K. & STAVRAKAKIS, G. S. 2007. Reinforcement learning for energy conservation and comfort in buildings. *Building and Environment*, 42, pp. 2686-2698.

DAS, S. 2001. Filters, wrappers and a boosting-based hybrid for feature selection. *In: Proceedings of the Eighteenth International Conference on Machine Learning*,28-June- 01 July, Williamstown, MA, USA, pp. 74-81.

DASARATHY, B. V. 1997. Sensor fusion potential exploitation-innovative architectures and illustrative applications. *Proceedings of the IEEE*, 85, pp. 24-38.

DASH, M., LIU, H. & MOTODA, H. 2000. Consistency based feature selection. *In: Proceedings of the Fourth Pacific Asia Conference on Knowledge Discovery and Data Mining*. 18-20 April, Kyoto, Japan, pp. 98-109.

DAVIDSSON, P. & BOMAN, M. 2005. Distributed monitoring and control of office buildings by embedded agents. *Information Sciences*, 171, pp. 293-307.

DE JONG, K. A. 1975. *An analysis of the behaviour of a class genetic adaptive systems.* Doctoral Dissertation, University of Michigan.

DEFRA. 2006. *e-Digest Statistics about :Climate Change* [Online]. Available: http://www.defra.gov.uk/environment/statistics/globatmos [Accessed 05/11/2010].

DEGELMAN, L. O. 1999. *A Model for Simulation of Daylighting and Occupancy Sensors as an Energy Control Strategy for Office Buildings* [Online]. Available: http://www.ibpsa.org/proceedings/BS1999/BS99_A-20.pdf [Accessed 06/08/2011].

DELANEY, D. T., O'HARE, G. M. & RUZZELLI, A. G. 2009. Evaluation of Energy-Efficiency in Lighting Systems Using Sensor Networks. *In: Proceedings of the First ACM Workshop on Embedded Sensing Systems for Energy Efficiency in Buildings*, 3 November , Berkeley,CA, USA, pp.19-24.

DELIS, A. L., DE CARVALHO, J. L. A., BORGES, G. A., DE SIQUEIRA RODRIGUES, S., DOS SANTOS, I. & DA ROCHA, A. F. 2009. Fusion of electromyographic signals with proprioceptive sensor data in myoelectric pattern recognition for control of active transfemoral leg prostheses. *In: Proceedings of Engineering in Medicine and Biology Society (EMBC 2009), Annual International Conference of the IEEE*, 3-6 Sept. 2009, Minnesota, USA, pp. 4755-4758.

DELNERO, C., HITTLE, D., ANDERSON, C., YOUNG, P., ANDERSON, M . 2001. Neural networks and PI control using steady state prediction applied to a heating coil. In: Proceedings of 7th REHVA World Congress and Clima 2000, 15-18 September, Naples, Italy, pp. 58-71.

DEVENDER & RAMASAMY, S. R. 1997. A review of EMI shielding and suppression materials. *In: Proceedings of the International Conference on Electromagnetic Interference and Compatibility*, 3-5 December, Hyderabad, India, pp. 459-466.

DEXTER, A. L., NGO,D., 2001. Fault Diagnosis in Air-Conditioning Systems : A Multi-Step Fuzzy Model Based Approach. *HVAC & R Research Journal*, 7, pp.83-102.

DING, C. & PENG, H. 2003. Minimum redundancy feature selection from microarray gene expression data. *In: Proceedings of the Computational Systems Bioinformatics conference*, 11-14 August , Stanford, California, pp. 523-529.

DOCTOR, F., HAGRAS, H. & CALLAGHAN, V. 2005. A fuzzy embedded agent-based approach for realizing ambient intelligence in intelligent inhabited environments. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 35, pp. 55-65.

DODIER, R. H., HENZE, G. P., TILLER, D. K. & GUO, X. 2006. Building occupancy detection through sensor belief networks. *Energy and Buildings*, 38, pp. 1033-1043.

DONG, B., ANDREWS, B., LAM, K. P., HÖYNCK, M., ZHANG, R., CHIOU, Y. & BENITEZ, D. 2010. An information technology enabled sustainability test-bed (ITEST) for occupancy detection through an environmental sensing network. *Energy and Buildings*, 42, pp.1038-1046.

DOUNIS, A. I. 2010. Artificial Intelligence for Energy Conversation in Buildings. *Advances in Building Energy Research*, 4, pp. 267-299.

DOUNIS, A. I., BRUANT, M., GUARRACINO, G., MICHEL, P. & SANTAMOURIS, M. 1996a. Indoor air-quality control by a fuzzy-reasoning machine in naturally ventilated buildings. *Applied Energy*, 54, pp. 11-28.

DOUNIS, A. I., BRUANT, M., SANTAMOURIS, M., GUARACCINO, G. & MICHEL, P. 1996b. Comparison of Conventional and Fuzzy Control of Indoor Air Quality in Buildings. *Journal of Intelligent and Fuzzy Systems*, 4, pp. 131-140.

DOUNIS, A. I. & CARAISCOS, C. 2009. Advanced control systems engineering for energy and comfort management in a building environment-A review. *Renewable and Sustainable Energy Reviews*, 13, pp. 1246-1261.

DTI 2002. Energy Consumption in the United Kingdom. (Department of Trade and Industry)

DUGELAY, J. L., JUNQUA, J. C., KOTROPOULOS, C., KUHN, R., PERRONNIN, F. & PITAS, I. 2002. Recent advances in biometric person authentication. *In:* 2002 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 13-17 May, Orlando, Florida, USA, pp.4060-4063.

EMMERICH, S. J. & PERSILY, A. K. 1997. Literature Review on CO2- Based Demand-Controlled Ventilation . *ASHRAE Transcations 103*,pp.4075.

EMMERICH, S. J. & PERSILY, A. K. 2001. *State-of-the-art review of CO2 demand controlled ventilation technology and application* [Online]. NISTIR 6729. Available: http://fire.nist.gov/bfrlpubs/build01/PDF/b01117.pdf [Accessed 05/04/2013].

EN15251:2007. Indoor Environmental Input Parameters for Design and Assessment of Energy Performance of Buildings Addressing Indoor Air Quality, Thermal Environment Lighting and Acoustics.

ENERGYPLUS. 2013a. *Input Output Reference* [Online]. Available: http://apps1.eere.energy.gov/buildings/energyplus/pdfs/inputoutputreference.pdf [Accessed 25/09/2013].

ENERGYPLUS. 2013b. *The Reference to EnergyPlus Calculations* [Online]. Available: http://apps1.eere.energy.gov/buildings/energyplus/pdfs/engineeringreference.pdf [Accessed 25/09/2013].

EPA. 1991. *Indoor Air Facts No. 4 (revised) Sick Building Syndrome*, U.S. Environmental Protection Agency. [Online]. Available: http://www.epa.gov/iaq/pdfs/sick_building_factsheet.pdf [Accessed 18/07/2013].

ERICKSON, V. L., CARREIRA-PERPINAN, M. A. & CERPA, A. E. 2011. OBSERVE: Occupancy-based system for efficient reduction of HVAC energy. *In: Proceedings of 10th International Conference on Information Processing in Sensor Networks (IPSN)*, 12-14 April, Chicago, IL, USA, pp. 258-269.

ERICKSON, V. L., LIN, Y., KAMTHE, A., BRAHME, R., SURANA, A. & CERPA, A. E. 2009. Energy Efficient Building Environment Control Strategies Using Real-time Occupancy Measurements. *Proceedings of the 1st ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings*, 4- 6 November, Berkeley, California, USA, pp. 19-24.

FAN, B., DU, Z., JIN, X., YANG, X. & GUO, Y. 2010. A hybrid FDD strategy for local system of AHU based on artificial neural network and wavelet analysis. *Building and Environment,* 45, pp. 2698-2708.

FISCHER, J. L. 2002. *Automatic meter reading module*. In: USPTO, editor. United States: Invensys Metering System/North America Inc. patent application.

FISK, W. J. 2008. *A pilot Study of the Accuracy of CO2 Sensors in Commercial Buildings* [Online]. Lawerence Berkeley National Laboratory. Available: http://escholarship.org/uc/item/78t0t90v [Accessed 20/12/2011].

FLOYD, D. B., PARKER, D. S., MCLLVAINE, J. E. R. & SHERWIN, J. R. 1995. *Energy efficiency technology demonstration projection for Florida educational facilities: Occupancy sensors* [Online]. Solar Energy Center Building Design Assistance Center. Available: http://eric.ed.gov/?id=ED433686 [Accessed 15/04/2012].

FOGARTY, C. A. & HUDSON, S. E. 2006. Sensing from the basement: A Feasibility Study of Unobtrusive and Low-cost Home Activity Recognition. *The nineteenth annual ACM Symposium on User Interface Software and Technology*, 15-18 October, Montreux, Switzerland, pp. 91-100.

FONTOYNONT, M., PLACE, W. & BAUMAN, F. 1984. Impact of electric lighting efficiency on the energy saving potential of daylighting from roof monitors. *Energy and Buildings,* 6, pp. 375-386.

FRISCHHOLZ, R. W. & DIECKMANN, U. 2000. BiolD: a multimodal biometric identification system. *Computer,* 33, pp. 64-68.

FUJISTU-SIEMENS. *Energy Savings with Personal Computers* [Online]. Available: http://www.fujistu-siemens.nl/aboutus/sor/energy_savings/prof_desk_prod.html [Accessed 20/11/2010].

FUNAKI, R., TANABE, S., TANAKA, H. & NAKAGAWA, T. 2003. Measurements of Chemical Emission Rates from Portable PC and Electronic Appliances. *Journal of Asian Architecture and Building Engineering*, 2, pp. 55-59.

GAKOVIC, B. 2000. Areas and types of glazing and other openings in the nondomestic building stock. *Environment and Planning B-Planning & Design*, 27, pp. 667-694.

GARG, V. & BANSAL, N. K. 2000. Smart occupancy sensors to reduce energy consumption. *Energy and Buildings*, 32, pp. 81-87.

GE-SENSING. *Telaire Ventostat Series* [Online]. Available: http://www.ge-mcs.com/en/co2/wall-mount/ventostat-8000-series.html [Accessed 11/02/2013].

GILLOTT, M., RODRIGUES, L. & SPATARU, C. 2010. Low-carbon Housing Design Informed by Research. *Proceedings of the Institution of Civil Engineers: Engineering Sustainability, 163*, pp. 77-87.

GILLOTT, M., SPATARU, C. & HALL, M. 2009. Domestic Energy and Occupancy : A Novel Post -Occupancy Evaluation Study. *8th International Conference on Sustainable Energy Technologies (SET2009)*, 31st August - 3rd September, Aachen, Germany.

GOLDBERG, D. E. 1989. *Genetic Algorithms in Search Optimization and Machine Learning*, Boston, MA, USA, Addison-Wesley Longman Publishing Co., Inc.

GOUDA, M. M., DANAHER, S., UNDERWOOD, C.P., 2006. Quasi-adaptive fuzzy heating control of solar buildings, *Building and Environment*, 41, pp. 1881–1891.

GUILLEMIN, A. & MOLTENI, S. 2002. An energy-efficient controller for shading devices self-adapting to the user wishes. *Building and Environment*, 37, pp.1091-1097.

GUILLEMIN, A. & MOREL, N. 2001. An innovative lighting controller integrated in a self-adaptive building control system. *Energy and Buildings*, 33, pp. 477-487.

GUYON, I. & ELISSEEFF, A. 2003. An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3, pp.1157–1182.

HAGRAS, H., PACKHARN, I., VANDERSTOCKT, Y., MCNULTY, N., VADHER, A. & DOCTOR, F. 2008. An intelligent agent based approach for energy management in commercial buildings. *In: Proceedings of IEEE International Conference on Fuzzy Systems, FUZZ-IEEE 2008. (IEEE World Congress on Computational Intelligence)*, 1-6 June, Hong Kong, pp.156-162.

HAILEMARIAM, E., GOLDSTEIN, R., ATTAR, R. & KHAN, A. 2011. Real-time occupancy detection using decision trees with multiple sensor types. *Proceedings of the Symposium on Simulation for Architecture and Urban Design*, 4-7 April, Boston, MA, pp. 141-148.

HALL, D. L. & LLINAS, J. 1997. An introduction to multisensor data fusion. *Proceedings of the IEEE*, 85, pp. 6-23.

HALL, D. L. & LLINAS, J. 2001. *Handbook of Multisensor Data Fusion*, CRC Press LLC.

HALL, D. L. & OGRODNIK. 1996. Passive Exploitation of the Electromagnetic Environment for Inmproved Target Tracking, Situation Assessment, and Threat Refinement. *In: Proceedings 9th National Symposuim on Sensor Fusion*, 12- 14 March, Monterey, Californa.

HALL, M. 1999. *Correlation based feature selection for machine learning.* Doctoral dissertation, University of Waikato.

HALL, M., FRANK, E., HOLMES, G., PFAHRINGER, B., REUTEMANN, P. & WITTEN, I. H. 2009. *The WEKA Mining Software: An Update*

HALL, M. A. & HOLMES, G. 2003. Benchmarking attribute selection techniques for discrete class data mining. *Knowledge and Data Engineering, IEEE Transactions on*, 15, pp. 1437-1447.

HASHEMIAN, H. M. 2005. *RTDs vs. thermocouples: Measuring industrial temperatures* [Online]. The Instrumentation, Systems, and Automation Society. Available: http://www.findarticles.com/p/articles/mi_qa3739/is_200309/ai_n9301173 [Accessed 22/01/2013].

HAZLEHURST, J. 2009. Chapter two - Industrial and Commercial Gas Meters. *Tolley's Industrial and Commercial Gas Installation Practice (Fifth edition).* Oxford: Newnes.

HE, X., CAI, D. & NIYOGI, P. 2005. Laplacian score for feature selection. *Advances in Neural Information Processing Systems*, 18, pp.507-514.

HEBB, D. 1949. *The organization of behaviour*, Wiley.

HENZE, G. P., DODIER, R.H., KRARTI,M., 1997. Development of a predictive optimal controller for thermal energy storage systems, *HVAC&R Research*, 3, pp. 233–264.

HERPEL, T., LAUER, C., GERMAN, R. & SALZBERGER, J. 2008. Multi-sensor data fusion in automotive applications. *In: Proceedings of 3rd International Conference on Sensing Technology, ICST 2008,* November 30 - December 3, Tainan, pp. 206-211.

HEWLETT-PARKARD. 2006. *2006 Global Citizen Report* [Online]. Available: http://www.hp.com/hpinfo/globalcitizenship/08gcreport/pdf/hp_fy06_gcr.pdf [Accessed 02/12/2010].

HOES, P., HENSEN, J. L. M., LOOMANS, M. G. L. C., DE VRIES, B. & BOURGEOIS, D. 2009. User behavior in whole building simulation. *Energy and Buildings*, 41, pp. 295-302.

HOLLAND, J. H. 1975. *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control and Artificial Intelligence.*, University of Michigan Press.

HONG, X., NUGENT, C., MULVENNA, M., MCCLEAN, S., SCOTNEY, B. & DEVLIN, S. 2009. Evidential fusion of sensor data for activity recognition in smart homes. *Pervasive and Mobile Computing*, 5, pp. 236-252.

HUANG, Y., LAN, Y., HOFFMANN, W. C. & LACEY, R. E. 2007. Multisensor Data Fusion for High Quality Data Analysis and Processing in Measurement and Instrumentation. *Journal of Bionic Engineering,* 4, pp. 53-62.

HUTCHINS, J., IHLER, A. & SMYTH, P. 2007. Modeling Count Data from Multiple Sensors: A Building Occupancy Model. *In: 2nd IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing, CAMPSAP 2007,* 12-14 December, St. Thomas, VI, pp. 241-244.

İÇOĞLU, O. & MAHDAVI, A. 2007. VIOLAS: A vision-based sensing system for sentient building models. *Automation in Construction,* 16, pp. 685-712.

IEA 1997. Demand Controlled Ventilating Systems: Summary of IEA Annex 18. International Energy Agency.

ION-SCIENCE. *TVOC Detector* [Online]. Available: http://www.ionscience.com/products?utm_source=BMONadwords&utm_medium=cpc&utm_content=Search%20Network&utm_campaign=Company [Accessed 12/02/2013].

IPCC. 2001. *Climate Change 2001: The Scientific Basis* [Online]. Intergovernmental Panel on Climate Change. Available: http://www.ipcc-wg2.org/ [Accessed 25/09/2011].

JAIN, L. C. 1989. Thermistor-based linear temperature-to-voltage converter. *Measurement,* 7, pp. 132-133.

JANG, J. S. R. 1993. ANFIS: adaptive-network-based fuzzy inference system. *Systems, Man and Cybernetics, IEEE Transactions on,* 23, pp. 665-685.

JARADAT, M. A. K. & LANGARI, R. 2009. A hybrid intelligent system for fault detection and sensor fusion. *Applied Soft Computing,* 9, pp. 415-422.

JIAN, L. & RUXU, D. 2005. Thermal comfort control based on neural network for HVAC application. *In: Proceedings of 2005 IEEE Conference on Control Applications, 2005. CCA 2005,* 28-31 August , Toronto, Canada, pp. 819-824.

JIANFENG, C., JIANMIN, Z., KAM, A. H. & SHUE, L. 2005. An automatic acoustic bathroom monitoring system. *In: IEEE International Symposium on Circuits and Systems, ISCAS 2005*, 23-26 May, pp. 1750-1753 Vol.2.

JINGWEN, T., MEIJUAN, G., HAO, Z. & KAI, L. 2007. Corrosion Detection System for Oil Pipelines Based on Multi-sensor Data Fusion by Wavelet Neural Network. *In: IEEE International Conference on Control and Automation, ICCA 2007,* May 30 2007-June 1, Guangzhou, China, pp. 2958-2963.

JOHN, G. H., KOHAVI, R. & PFLEGER, K. 1994. Irrelevant feature and the subset selection problem. *In: COHEN, W. W., AND HIRSH, H., ed. Machine Learning: Proceedings of the Eleventh International Conference*, New Brunswick, N.J., Rutgers University. pp. 121-129.

KAR, S. & VARSHNEY, P. K. 2009. Accurate estimation of indoor occupancy using gas sensors. *In: 5th International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP),* 7-10 Decemeber, pp. 355-360, Melbourne, VIC.

KATIPAMULA, S. & BRAMBLEY, M. R. 2005. Methods for fault detection, diagnostics and prognostics for building systems—a review. *IHVAC&R Research: Part I*, 11, pp. 3-25.
KAUSHIK, A. R. & CELLER, B. G. 2007. Characterization of PIR Detector for Monitoring Occupancy Patterns and Functional Health Status of Elderly People Living Alone at Home. *Technology and Health Care*, 15, pp. 273-288.
KAWAMOTO, K., SHIMODA, Y. & MIZUNO, M. 2004. Energy saving potential of office equipment power management. *Energy and Buildings*, 36, pp. 915-923.
KAYA, I., TAN, N., ATHERTON,D.P., 2007. Improved cascade control structure for enhanced performance. *Journal Process Control*, 17, pp. 3-16.
KIM, Y., SCHMID, T., CHARBIWALA, Z. M. & SRIVASTAVA, M. B. 2009. ViridiScope: Design and Implementation of a Fine Grained Power Monitoring System for Homes. *Proceedings of the 11th international conference on Ubiquitous computing*, 30 September - 03 October, Orlando, Florida, pp. 245-254.
KIRA, K., AND RENDELL, L.A. 1992 . A practical approach to feature selection. *In: SLEEMAN AND EDWARDS, P., ed. Proceedings of the Ninth International Conference on Machine Learning (ICML-92)*, pp. 249-256.
KLEIN, L., KWAK, J.-Y., KAVULYA, G., JAZIZADEH, F., BECERIK-GERBER, B., VARAKANTHAM, P. & TAMBE, M. 2012. Coordinating occupant behavior for building energy and comfort management using multi-agent systems. *Automation in Construction*, 22, pp. 525-536.
KOHAVI, R. & JOHN, G. 1997. Wrappers for Feature Subset Selection. *Artificial Intelligence*, pp. 273-324.
KOLOKOTSA, D. 2007. Artificial Intelligence in Buildings: A Review of the Application of Fuzzy Logic. *Advances in Building Energy Research*, 1, pp. 27-54.
KOON, W. 2002. *Current Sensing for Energy Metering* [Online]. Shanghai, China: Analog Devices. Available: http://www.analog.com/static/imported-files/tech_articles/16174506155607IIC_Paper.pdf [Accessed 10/10/2013].
KORHONEN, I., PARKKA, J. & VAN GILS, M. 2003. Health monitoring in the home of the future. *Engineering in Medicine and Biology Magazine, IEEE*, 22, pp. 66-73.
KORN, D., HUANG, R., BEAVERS, D., BOLIOLI, T. & WALKER, M. 2004. Power management of computers. *In: IEEE International Symposium on Electronics and the Environment*, 10-13 May, pp. 128-131.
KORN, D., HUANG, R., BOLIOLI, T. & WALKER, M. 2006. Computer Power Management for Enterprises A Practical Guide for Saving up to $100 per Seat Annually in Electricity. *In: Proceedings of the 2006 IEEE International Symposium on Electronics and the Environment*, 8-11 May, pp.161-166.
KRARTI, M. 2003. *An overview of Artificial Intelligence-Based Methods for Building Energy Systems* [Online]. Available: http://scitation.aip.org/getpdf/servlet/GetPDFServlet?filetype=pdf&id=JSEEDO000125000003000331000001&idtype=cvips&prog=normal [Accessed 18/07/2012].

KRUMM, J., ABOWD, G., SENEVIRATNE, A., STRANG, T., PATEL, S., ROBERTSON, T., KIENTZ, J. & REYNOLDS, M. 2007. At the Flick of a Switch: Detecting and Classifying Unique Electrical Events on the Residential Power Line, *In: Proceedings of the 9th international conference on Ubiquitous computing. UbiComp 200,* Springer Berlin / Heidelberg, pp. 271-288.

KUKOLJ, D. D., KUZMANOVIC, S.B., LEVI, E., 2001. Design of a PID-like compound fuzzy logic controller. *Engineering Applications of Artificial Intelligence*, 14, pp. 785–803.

KURIAN, C. P., KURIACHAN, S., BHAT, J. & AITHAL, R. S. 2005. An Adaptive Neuro-Fuzzy Model for the Prediction and Control of Light in Integrated Lighting Scheme. *Lighting Research and Techology*, 37, pp. 343-352.

KUSIAK, A., LI, M. & ZHENG, H. 2010. Virtual models of indoor-air-quality sensors. *Applied Energy*, 87, pp. 2087-2094.

LAM, J. C. & LI, D. H. W. 2003. Electricity consumption characteristics in shopping malls in subtropical climates. *Energy Conversion and Management,* 44, pp. 1391-1398.

LAM, K. P., HOYNCK, M., DONG, B., ANDREWS, B., CHIOU, Y. & ZHANG, R. 2009a. Occupancy detection through an extensive environmental sensork network in an open-plan office building. *Proceedings of the Eleventh International IBPSA conference,* 27-30 July, Glasgow, Scotland, pp. 1452-1459.

LAM, K. P., HOYNCK, M., ZHANG, R., ANDREWS, B., CHIOU, Y., DONG, B. & BENITEZ, D. 2009b. Information-Theoretic Environmental Features Selection for Occupancy Dtection in Open Offices. *Eleventh International IBPSA Conference.* 27-30 July, Glasgow, Scotland, pp. 1460-1467.

LAM, N. 1995. Intelligent computer control of air conditioning system based on genetic algorithm and classifier system. *In: The Proceedings of the 1995 Building Simulation Conference,* August, Wisconsin, USA, pp. 151-157.

LEE, H.-Y., YANG, I. T. & LIN, Y.-C. 2012. Laying out the occupant flows in public buildings for operating efficiency. *Building and Environment*, 51, pp. 231-242.

LEVERMORE, G. J. 2000. *Building Energy Management Systems in Applications to Low Energy HVAC and Natural Ventilation*, E & FN Spon.

LI, D. H. W. & LAM, J. C. 2003. An investigation of daylighting performance and energy saving in a daylit corridor. *Energy and Buildings,* 35, pp. 365-373.

LI, H., AND BRAUN, J.E., 2009. Virtual refrigerant pressure sensors for use in monitoring and fault diagnosis of vaporcompression equipment. *HVAC&R Research* 15, pp. 597-616.

LI, H. & BRAUN, J. E. 2007. Decoupling features and virtual sensors for diagnosis of faults in vapor compression air conditioners. *International Journal of Refrigeration* 30, pp. 546-564.

LI, H., YU, D. & BRAUN, J. E. 2011. A review of virtual sensing technology and application in building systems. *HVAC & R Research Journal,* 17, pp. 619-645.

LI, N., CALIS, G. & BECERIK-GERBER, B. 2012. Measuring and monitoring occupancy with an RFID based system for demand-driven HVAC operations. *Automation in Construction*, 24, pp. 89-99.

LI, P. Y., CHEN, M. S., HIBINO, H., KOYAMA, S. & ZHENG, M. C. 2009a. Rest facilities at commercial plazas through user behavior perspective. *Journal of Asian Architecture and Building Engineering*, pp. 127–134.

LI, Q., MENG, Q., CAI, J., YOSHINO, H. & MOCHIDA, A. 2009b. Predicting hourly cooling load in the building: A comparison of support vector machine and different artificial neural networks. *Energy Conversion and Management,* 50, pp. 90-96.

LI, W. M., LEE, S. C. & CHAN, L. Y. 2001. Indoor air quality at nine shopping malls in Hong Kong. *Science of The Total Environment*, 273, pp. 27-40.

LI, Y., ANG, K.H., CHONG, G.C.Y., 2006. PID control system analysis and design – problems, remedies, and future directions,. *IEEE Control Systems Magazine.*

LIAO, C., LIN, Y. & BAROOAH, P. 2012. Agent-based and graphical modelling of building occupancy. *Journal of Building Performance Simulation,* 5, pp. 5-25.

LINDEN, P. F. 1999. The Fluid Mechanics of Natural Ventilation. *Annual Review of Fluid Mechanics* 31, pp. 201-238.

LIU, H., AND MOTODA, H., 1998. *Feature selection for knowledge discovery and data mining*, Boston: Kluwer Academic Publishers.

LÓPEZ-RODRÍGUEZ, M. A., SANTIAGO, I., TRILLO-MONTERO, D., TORRITI, J. & MORENO-MUNOZ, A. 2013. Analysis and modeling of active occupancy of the residential sector in Spain: An indicator of residential electricity consumption. *Energy Policy,* 62, pp. 742-751.

LÖTJÖNEN, J., KORHONEN,I., HIRVONEN,K., ESKELINEN,S., MYLLYMÄKI,M., PARTINEN,M., 2003. Automatic sleep/wake and nap analysis with a new wrist worn online activity monitoring device Vivago WristCare. *Sleep,* 26, pp. 86-90.

LOVEDAY, D. L. & VIRK, G. S. 1992. Artificial Intelligence for Buildings. *Applied Energy,* 41, pp. 201-221.

LU, J., SOOKOOR, T., SRINIVASAN, G. G., HOLBEN, B., STANKOVIC, E., FIELD, J. & WHITEHOUSE, K. 2010. The smart thermostat: using occupancy sensors to save energy in homes. *In :* The 8th ACM Conference on Embedded Networked Sensor Systems, *SenSys, 2010,* 3-5 November, Zurich, Switzerland.

LYNNWORTH, L. C. & LIU, Y. 2006. Ultrasonic flowmeters: Half-century progress report, 1955–2005. *Ultrasonics,* 44, pp. 1371-1378.

MAINALI, K. & ORUGANTI, R. 2010. Conducted EMI Mitigation Techniques for Switch-Mode Power Converters: A Survey. *Power Electronics, IEEE Transactions on,* 25, pp. 2344-2356.

MAMIDI, S., CHANG, Y. & MAHESWARAN, R. 2012. Improving Building Energy Efficiency with a Network of Sensing, Learning and Prediction Agents. *In: Proceedings of the 11th International Conference on*

Autonomous Agents and Multiagent Systems, 4-8th June, Valencia, Spain, pp. 45-52 .

MANICCIA, D. & WOLSEY, R. 1998. *Occupancy Sensors* [Online]. National Lighting Product Information Program. Available: http://www.lrc.rpi.edu/nlpip/publicationDetails.asp?id=102&type=1 [Accessed 10/05/2012].

MARTANI, C., LEE, D., ROBINSON, P., BRITTER, R. & RATTI, C. 2012. ENERNET: Studying the dynamic relationship between building occupancy and energy consumption. *Energy and Buildings*, 47, pp. 584-591.

MARTINS, F. G., COELHO, M.A.N., 2000. Application of feedforward artificial neural networks to improve process control of PID-based control algorithms. *Computers and Chemical Engineering*, 24, pp. 853–858.

MELFI, R., ROSENBLUM, B., NORDMAN, B. & CHRISTENSEN, K. 2011. Measuring building occupancy using existing network infrastructure. *In: Proceedings of 2011 IEEE International Green Computing Conference*,25-28 July, Orlando, Florida, pp. 1-8.

MEYERS, R. J., WILLIAMS, E. D. & MATTHEWS, H. S. 2010. Scoping the potential of monitoring and control technologies to reduce energy use in homes. *Energy and Buildings*, 42, pp. 563-569.

MEYN, S., SURANA, A., YIQING, L., OGGIANU, S. M., NARAYANAN, S. & FREWEN, T. A. 2009. A sensor-utility-network method for estimation of occupancy in buildings. *In: Proceedings of the 48th IEEE Conference on Decision and Control, 2009 held jointly with the 28th Chinese Control Conference, CDC/CCC 2009*, 15-18 December, Shanghai, China, pp. 1494-1500.

MILLS, E. 1993. Efficient Lighting Programs in Europe - Cost-Effectiveness, Consumer Response, and Market Dynamics, *Energy*, 18, pp. 131-144.

MOGHAVVEMI, M. & Seng, L.C . 2004. Pyroelectric Infrared sensor for Intruder Detection. *In Proceedings of IEEE Region 10 Conference on Analog and Digital Techniques in Electrical Engineering, TENCON 2004,* 21-24 November, Chiang Mai, Thailand, pp. 656-659.

MORENO-MUNOZ, A., FLORES-ARIAS, J. M., GIL-DE-CASTRO, A. & DE LA ROSA, J. G. 2009. Power quality and energy efficiency in e-offices. *In: 35th Annual Conference of IEEE Industrial Electronics, IECON '09*, 3-5 November, Porto, Portugal, pp. 748-752.

MOSHNYAGA, V. G. 2008. How to Really Save Computer Energy? *Proceedings on the International Conference on Computer Design (CDES 2008), 14-17 July*, Las Vegas, USA, pp. 89-95.

MOSHNYAGA, V. G. 2010. A new energy management approach for user-centric applications. *In: 2010 International Conference on Green Circuits and Systems (ICGCS)* , 21-23 June, Shanghai, China, pp. 1-6 .

MOTUZIENE, V. & VILUTIENE, T. 2013. Modelling the Effect of the Domestic Occupancy Profiles on Predicted Energy Demand of the Energy Efficient House. *Procedia Engineering*, 57, pp. 798-807.

MUNGWITITKUL, W. & MOHANTY, B. 1997. Energy efficiency of office equipment in commercial buildings: The case of Thailand. *Energy*, 22, pp. 673-680.

NAKAYAMA, H., ANSARI, N., JAMALIPOUR, A. & KATO, N. 2007. Fault-resilient sensing in wireless sensor networks. *Computer Communications,* 30, pp. 2375-2384.

NASSAR, K. 2010. A model for assessing occupant flow in building spaces. *Automation in Construction,* 19, pp.1027-1036.

NASSIF, N., MOUJAES, S. & ZAHEERUDDIN, M. 2008. Self-tuning dynamic models of HVAC system components. *Energy and Buildings,* 40, pp.1709-1720.

NGO, D., DEXTER, A.L., 1999. A Robust Model Based Approach to Diagnosing Faults in Air- Handling Units *ASHRAE Trans.,* 105, pp.1078-1086.

NGUYEN, T. A. & AIELLO, M. 2013. Energy intelligent buildings based on user activity: A survey. *Energy and Buildings,* 56, pp.244-257.

NIELSEN, T. R. & DRIVSHOLM, C. 2010. Energy efficient demand controlled ventilation in single family houses. *Energy and Buildings,* 42, pp.1995-1998.

O'DRISCOLL, E. & O'DONNELL, G. E. 2013. Industrial power and energy metering – a state-of-the-art review. *Journal of Cleaner Production,* 41, pp. 53-64.

O’DRISCOLL, E., CUSACK, D. O. & O’DONNELL, G. E. 2012. Implementation of Energy Metering Systems in Complex Manufacturing Facilities–A Case Study in a Biomedical Facility. *Procedia CIRP,* 1, pp. 524-529.

OGC. 2008. *ICT Power Management Project Case Study* [Online]. Available: www.ogc.gov.uk [Accessed 28/ 11/2010].

OKUN, O. 2011. *Feature Selection and Ensemble Methods for Bioinformatics: Algorithmic Classification and Implementations*, Medical Information Science Reference (IGI Global).

ONSET-CORPERATION. *HOBO U-Series Loggers* [Online]. Available: http://www.onsetcomp.com/ [Accessed 15/05/2012].

PADMANABH, A., AMRIT, S., VUPPALA, S., MALIKARJUNA, V., KUMAR, S. & PAUL, S. 2009. A Wireless Sensor Network Based Conference Room Management System. *In BuildSys ’09: Proceedings of the First ACMWorkshop on Embedded Sensing Systems for Energy-Efficiency in Buildings*, 4- 6 November, Berkeley, CA, USA, pp. 37 -42.

PAGE, J., ROBINSON, D., MOREL, N. & SCARTEZZINI, J. L. 2008. A generalised stochastic model for the simulation of occupant presence. *Energy and Buildings,* 40, pp. 83-98.

PAINTER, B., BROWN, N. & COOK, M. J. 2012. Practical application of a sensor overlay system for building monitoring and commissioning. *Energy and Buildings,* 48, pp.29-39.

PARGFRIEDER, J. & JORGL, H. P. 2002. An integrated control system for optimizing the energy consumption and user comfort in buildings. *In: Proceedings of 2002 IEEE International Symposium on Computer Aided Control System Design*, pp. 127-132.

PARK, J. S. & KIM, H. J. 2012. A field study of occupant behavior and energy consumption in apartments with mechanical ventilation. *Energy and Buildings,* 50, pp.19-25.

PARYS, W., SAELENS, D. & HENS, H. 2011. Coupling of dynamic building simulation with stochastic modeling of occupant behaviour in offices – a

reviewbased integrated methodology. *Journal of Building Performance Simulation*, 1, pp.1-20.
PAVLOVAS, V. 2004. Demand controlled ventilation: A case study for existing Swedish multifamily buildings. *Energy and Buildings*, 36, pp.1029-1034.
PAYNE, R. K. & LIEN, W. A. 2011. *Automated meter reader direct mount endpoint module*. In: USPTO, editor. United States: Itron, Inc. (Liberty Lake, WA, US) patent application.
PEJCIC, B., EADINGTON, P., Ross, A . 2007. Environmental monitoring of hydrocarbons: a chemical sensor perspective . *Environmental Science & Technology*, 41, pp.6333-6342.
PÉREZ-LOMBARD, L., ORTIZ, J. & POUT, C. 2008. A review on buildings energy consumption information. *Energy and Buildings*, 40, pp.394-398.
PESCHIERA, G., TAYLOR, J. E. & SIEGEL, J. A. 2010. Response–relapse patterns of building occupant electricity consumption following exposure to personal, contextualized and occupant peer network utilization data. *Energy and Buildings*, 42, 1329-1336.
PIATETSKY-SHAPIRO, G. 2005. *KDnuggets news on SIGKDD service award.* [Online]. Available: http://www.kdnuggets.com/news/ [Accessed 20/04/2011].

POTTER, I. & BOOTH, W. B. 1994. CO2 controlled mechanical ventilation systems. *In:* BSRIALTD (ed.) *Technical note TN 12/94.1 (BSRIA).* Bracknell, UK, .
RAFIQ, M. Y., BUGMANN, G. & EASTERBROOK, D. J. 2001. Neural network design for engineering applications. *Computers & Structures*, 79, pp. 1541-1552.
RAJ, P. A., SUDHAKARAN, M. & RAJ, P. P. 2009. Estimation of Stand-by Power Consumption for Typical Appliances *Journal of Engineering Science and Technology Review*, 2, pp. 71-75.
RCEP. 2000. *Energy – The Changing Climate* [Online]. Royal Commission on Environmental Pollution. Available: http://www.rcep.org.uk/newenergy.htm [Accessed 20/09/2011].
REINHART, C. F. 2004. Lightswitch-2002: a model for manual and automated control of electric lighting and blinds. *Solar Energy*, pp. 77, 15-28.
RICHARDSON, I., THOMSON, M. & INFIELD, D. 2008. A high-resolution domestic building occupancy model for energy demand simulations. *Energy and Buildings*, 40, pp. 1560-1566.
RICHARDSON, I., THOMSON, M., INFIELD, D. & CLIFFORD, C. 2010. Domestic electricity use: A high-resolution energy demand model. *Energy and Buildings*, 42, pp. 1878-1887.
ROBERSON, J. A., BROWN, R. E., NORDMAN, B., WEBBER, C. A., HOMAN, G. K., MAHAJAN, A., MCWHINNEY, M. & KOOMEY, J. G. 2002. *Power Levels in Office Equipment:Measurements of New Monitors and Personal Computers* [Online]. Available: http://www.osti.gov/bridge/purl.cover.jsp?purl=/799608-nlts28/native/ [Accessed 02/11/2010].

ROSENBLATT, F. 1958. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological review*, 65, pp.386–408.

ROUF, I., MUSTAFA, H., XU, M., XU, W., MILLER, R. & GRUTESER, M. 2012. Neighborhood watch: security and privacy analysis of automatic meter reading systems. *In: Proceedings of the 2012 ACM conference on computer and communications* , 16- 18 October, Raleigh, North Carolina, USA, pp. 462–73.

ROVETI, D. 2005. *Choosing a Humidity Sensor: A Review of Three Technologies* [Online]. Available: http://www.sensorsmag.com/articles/0701/54/main.shtml [Accessed 25/06/2012].

RUMELHART, D., HINTON, G. & WILLIAMS, R. 1986. Learning representations by back-propagating errors. *Nature*, 323, pp. 533–536.

RUNQUIST, R., MCDOUGAL, T. & BENYA, J. 1996. *Lighting Controls : Patterns for Design* [Online]. CA: The Electric Power Research Institute. Available: http://www.lightingassociates.org/i/u/2127806/f/tech_sheets/Lighting_Controls_Patterns_for_Design.pdf [Accessed 17/01/2012].

RUTISHAUSER, U., JOLLER, J. & DOUGLAS, R. 2005. Control and learning of ambience by an intelligent building. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 35, pp.121-132.

S+S. *Aerasgard rlq-series air quality sensor* [Online]. Available: http://www.spluss.eu/5-air-quality-sensor-flow/16-air-quality-sensor-co2-sensor/61-room-air-quality-sensor-voc-on-wall/117-rlq/en [Accessed 06/08/2013].

SANDERSON, M. L. & YEUNG, H. 2002. Guidelines for the use of ultrasonic non-invasive metering techniques. *Flow Measurement and Instrumentation*, 13, pp. 125-142

SARKAR, A., FAIRCHILD, M. & SALVAGGIO, C. 2008. Integrated daylight harvesting and occupancy detection using digital imaging. *Proceedings for Sensors, Cameras and System for Industrial/Scientific Applications , IX, SPIE.* San Jose, CA, USA.

SCHELL, M. 2008. *Making Sense out of Sensors* [Online]. AirTest Technologies Inc. Available: https://www.airtest.com/support/reference/article1.pdf [Accessed 08/10/2011].

SCHELL, M., INTHOUT, D., 2001. Demand control ventilation using CO2. *ASHRAE Journal*, 2, pp. 1-6.

SEE, L. & ABRAHART, R. J. 2001. Multi-model data fusion for hydrological forecasting. *Computers & Geosciences*, 27, pp. 987-994.

SENTHAMARAI KANNAN, S., RAMARAJ, N., 2010. A novel hybrid feature selection via Symmetrical Uncertainty ranking based local memetic search algorithm. *Knowledge-Based Systems*, 23, pp. 580-585.

SHIRI, J. & KIŞI, Ö. 2011. Comparison of genetic programming with neuro-fuzzy systems for predicting short-term water table depth fluctuations. *Computers & Geosciences*, 37, pp. 1692-1701.

SHRESTHA, S. & MAXWELL, G. 2010. *Product Testing Report Supplement :Wall Mounted Carbon Dioxide (CO2) Transmitters* [Online]. National Buildings

Controls Information Program. Available: http://www.energy.iastate.edu/Efficiency/Commercial/download_nbcip/PTR_CO2_3_2010SUPPfin.pdf [Accessed 20/03/2012].
SIERRA, E. G.-M., R; HOSSIAN, A; BRITOS, P; BALBUENA, E; 2006. Providing Intelligent User-Adapted Control Strategies in Building Environment. *Research in Computing Science* 19, pp. 235-241.
SILVESTRE, B. J. & PÉREZ, L., R., 2011. Energy efficiency improvements through surveillance applications in industrial buildings. *Energy and Buildings*, 43, pp. 1334-1340.
SOKWOO, R., BOO-HO, Y., KUOWEI, C. & ASADA, H. H. 1998. The ring sensor: a new ambulatory wearable sensor for twenty-four hour patient monitoring. *In: Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 29 October -1 November , Hong Kong, pp. 1906-1909 vol.4.
SOMAVAT, P. J., S; NAMBOODIRI, V. 2010. Accounting for the Energy Comsumption of Personal Computing Including Portable Devices. *Proceedings of the 1st International Conference on Energy-Efficient Computing and Networking*, 13-15 April, Passua, Germany, pp. 141-149.
STOKES, M., RYLATT, M. & LOMAS, K. 2004. A simple model of domestic lighting demand. *Energy and Buildings*, 36, pp.103-116.
SUN, Z., WANG, S. & MA, Z. 2011. In-situ implementation and validation of a CO2-based adaptive demand-controlled ventilation strategy in a multi-zone office building. *Building and Environment*, 46, pp. 124-133.
SUNGMEE, P. & JAYARAMAN, S. 2003. Enhancing the quality of life through wearable technology. *Engineering in Medicine and Biology Magazine, IEEE*, 22, pp.41-48.
TABAK, V. & DE VRIES, B. 2010. Methods for the prediction of intermediate activities by office occupants. *Building and Environment*, 45, pp. 1366-1372.
TACHWALI, Y., REFAI, H. & FAGAN, J. E. 2007. Minimizing HVAC Energy Consumption Using a Wireless Sensor Network. *In:* 33rd Annual Conference of the IEEE Industrial Electronics Society, IECON 2007., 5-8 November, pp. 439-444.
TARZIA, S. P., DICK, R. P., DINDA, P. A. & MEMIK, G. 2009. Sonar-based measurement of user presence and attention. *In Ubicomp '09: Proceedings of the 11th international conference on Ubiquitous computing, ACM*, 30 September - 03 October, New York, NY, USA, pp.89-92.
TEWOLDE, M., LONGTIN, J. P., DAS, S. R. & SHARMA, S. 2013. Determining appliance energy usage with a high-resolution metering system for residential natural gas meters. *Applied Energy*, 108, pp. 363-372.
THOMAS, B., SOLEIMANI-MOHSENI, M. & FAHLÉN, P. 2005. Feed-forward in temperature control of buildings. *Energy and Buildings*, 37, pp. 755-761.
TILLER, D. K., GUO, X., HENZE, G. P. & WATERS, C. E. 2010. Validating the Application of Occupancy Sensor Networks for Lighting Control. *Lighting Research and Techology*, 42, pp.399-414.
TOMASTIK, R., YIQING, L. & BANASZUK, A. 2008. Video-based estimation of building occupancy during emergency egress. *In: American Control Conference*,11-13 June, Seattle, USA, pp. 894-901.

TORRITI, J. 2012. Demand Side Management for the European Supergrid: Occupancy variances of European single-person households. *Energy Policy,* 44, pp.199-206.
TSI. *ALNOR LOFLO BALOMETER CAPTURE HOOD 6200D* [Online]. TSI-ALNOR USA. Available: http://www.tsi.com/alnor-loflo-balometer-capture-hood-6200d/ [Accessed 10/10/2013].
TURNER, C. & FRANKEL, M. 2008. *Energy Performance of LEED for New Construction Buildings* [Online]. U.S. Green Building Council. Available: https://wiki.umn.edu/pub/PA5721_Building_Policy/WebHome/LEEDENERGYSTAR_STUDY.pdf [Accessed 08/08/2013].
VARIOUS. 2005. *Digital CFM Airflow Meter – CFM Master II* [Online]. Available: http://www.terrauniversal.com/products/measuring/digitalcfm.php [Accessed 09/10/2011].
VITAMINCM. 2008. *Reusing an Old PC as a Server* [Online]. Available: http://www.vitamincm.com/wp-content/uploads/2008/01/reuse-an-old-pc.pdf [Accessed 19/12/2010].
WALKER, J. M. 2009. Power management for networked computers: A review of incentive programs. *In: IEEE International Symposium on Sustainable Systems and Technology, 2009, ISSST '09*, 18-20 May, Phoenix, Arizona, USA, pp.1-6.
WALTZ, E. 1995. The Principles and Practice of Image and Spatial Data Fusion. *Proceedings of 8th National Data Fusion Conference.* 15-17 March , Dallas, Texas, pp. 257-278.
WANG, B.-L., TAKIGAWA, T., YAMASAKI, Y., SAKANO, N., WANG, D.-H. & OGINO, K. 2008. Symptom definitions for SBS (sick building syndrome) in residential dwellings. *International Journal of Hygiene and Environmental Health,* 211, pp. 114-120.
WANG, D., FEDERSPIEL, C. C. & RUBINSTEIN, F. 2005. Modeling occupancy in single person offices. *Energy and Buildings,* 37, pp. 121-126.
WANG, S. & JIN, X. 1998. CO2 - Based Occupancy Detection for On-line Outdoor AirFlow Control. *Indoor and Built Environment,* 7, pp. 165-181.
WANG, W.C., CHAU, K.W., CHENG, C.T. & QIU, L. 2009. A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series. *Journal of Hydrology,* 374, pp. 294-306.
WARREN, P. 2014. A review of demand-side management policy in the UK. *Renewable and Sustainable Energy Reviews,* 29, pp. 941-951.
WBCSD 2008. Energy Efficiency in Buildings Brief #2- Our Vision: A World Buildings Consume Zero Net Energy. 24/09/2011 ed.: World Business Council for Sustainable Development.
WEBBER, C. A., ROBERSON, J. A., MCWHINNEY, M. C., BROWN, R. E., PINCKARD, M. J. & BUSCH, J. F. 2006. After-hours power status of office equipment in the USA. *Energy,* 31, pp. 2823-2838.
WEBBER, L. 2007. *Computer Use Expected to Top 2 Billion* [Online]. Available: http://www.inc.com/news/articles/200707/computers.html [Accessed 20/10/2010].
WHITEHOUSE, K., RANJAN, J., LU, J., SOOKOOR, T., SAADAT, M., BURKE, C. M., STAENGL, G., CANFORA, A. & HAJ-HARIRI, H. 2012. Towards

Occupancy-Driven Heating and Cooling. *Design & Test of Computers, IEEE,* 29, pp. 17-25.

WILSON, B. 1998. *The Machine Learning Dictionary* [Online]. Available: http://www.cse.unsw.edu.au/~billw/dictionaries/mldict.html [Accessed 04/10/2013].

WILSON, D. H. & ATKESON, C. G. 2005. Simultaneous Tracking and Activity Recognition (STAR) Using Many Anonymous, Binary Sensors. *In: Proceedings of the International Conference on Pervasive Computing (Pervasive 2005)*, 8-13 May, Munich, Germany, pp.62-79.

WITTEN, I. H. & FRANK, E. 2000. *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations,* San Francisco,, Morgan Kaufmann,.

WOLKOFF, P. & KJÆRGAARD, S. K. 2007. The dichotomy of relative humidity on indoor air quality. *Environment International,* 33, pp. 850-857.

WOLKOFF, P., NIELSEN, G.D., 2001. Organic compounds in indoor air – their relevance for perceived indoor air quality. *Atmospheric Environment,* 35, pp. 4407–4417.

WON, D. & YANG, W. 2005. *The State of -the-Art in Sensor Technology for Demand-Controlled Ventilation* [Online]. Canada: Institute for Research in Construction. Available: http://www.nrc-cnrc.gc.ca/obj/irc/doc/pubs/rr/rr243/rr243.pdf [Accessed 18/06/2011].

WRIGHT, J. A., LOOSEMORE, H. A. & FARMANI, R. 2002. Optimization of building thermal design and control by multi-criterion genetic algorithm. *Energy and Buildings,* 34, pp. 959-972.

WU, X., DE-AN, Z., JINLIANG, G. & BO, C. 2010. Research on the Abrasive Water-Jet Cutting Machine Information Fusion Fault Diagnosis System Based on Fuzzy Neural Network. *In: 2010 International Conference on Biomedical Engineering and Computer Science (ICBECS)*, 23-25 April, Wuhan, pp. 1-4.

WUSHENG, W., MENGFEN, H. & CHUNGLIN, H. 2008. People tracking and counting for applications in video surveillance system. *In:International Conference on Audio, Language and Image Processing, ICALIP 2008, 7-9 July,* Shanghai, pp.1677-1682.

YANG, R. & WANG, L. 2013. Development of multi-agent system for building energy and comfort management based on occupant behaviors. *Energy and Buildings*, 56, pp.1-7.

YANG, X.B., JIN, X.Q., DU, Z.M., FAN, B. & CHAI, X.F. 2011. Evaluation of four control strategies for building VAV air-conditioning systems. *Energy and Buildings*, 43, pp.414-422.

YANG, Z., LI, N., BECERIK-GERBER, B. & AND OROSZ, M. 2012. A Non-Intrusive Occupancy Monitoring System for Demand Driven HVAC Operations. *Construction Research Congress 2012.* American Society of Civil Engineers.

YIFEI, C., HUAI, L. & XUELIANG, C. 2009. Venetian Blind Control System Based on Fuzzy Neural Network for Indoor Daylighting. *In: Second International Conference on Computer and Electrical Engineering, ICCEE '09*, 28-30 December, pp. 269-273, Dubai.

YU, L., LIU, H. & GUYON, I. 2004. Efficient feature selection via analysis of relevance and redundancy. *Journal of Machine Learning Research*, 5, pp.1205-1224.

YU, T. 2010. Modeling Occupancy Behavior for Energy Efficiency and Occupants Comfort Management in Intelligent Buildings. *In: 2010 Ninth International Conference on Machine Learning and Applications (ICMLA)*, 12-14 December, Washington, DC, USA, pp. 726-731.

YUDELSON, J. 2010. *Greening Existing Buildings*, New York, Green Source/McGraw-Hill.

ZADEH, L. 1965. Fuzzy sets. *Information and control*, 8, 338–353.

ZAHEER-UDDIN, M., ZHENG,G.R., 2000. Optimal control of time scheduled heating, ventilating and air conditioning processes in buildings. *Energy Conversion and Management* 41, pp.49-60.

ZAMPOLLI, S., ELMI, I., AHMED, F., PASSINI, M., CARDINALI, G. C., NICOLETTI, S. & DORI, L. 2004. An electronic nose based on solid state sensor arrays for low-cost indoor air quality monitoring applications. *Sensors and Actuators B: Chemical*, 101, pp. 39-46.

ZEILER, W. H., R., KAMPHUIS, R., HOMMELBERG, M., 2006. Agent Technology to Improve Building Energy Efficiency and Occupants Comfort. *Proceedings of the International Conference for Enhanced Building Operations.* Shenzhen, China.

ZERVAS, E., MPIMPOUDIS, A., ANAGNOSTOPOULOS, C., SEKKAS, O. & HADJIEFTHYMIADES, S. 2010. Multisensor data fusion for fire detection. *Information Fusion*, 12, pp.150 -159.

ZHAO, Z., MORSTATTER, F., SHARMA, S., ALELYANI, S., ANAND, A. & LIU, H. 2010. *Advancing Feature Selection Research- ASU Feature Selection Repository*[Online]. Available: http://featureselection.asu.edu/featureselection_techreport.pdf [Accessed 28/11/2012].

ZHENG, L., FORSYTH, D. S., KOMOROWSKI, J. P., HANASAKI, K. & KIRUBARAJAN, T. 2007. Survey: State of the Art in NDE Data Fusion Techniques. *Instrumentation and Measurement, IEEE Transactions on,* 56, pp. 2435-2451.

ZHU, W., RUI, Y. & LINGFENG, W. 2010. Multi-agent intelligent controller design for smart and sustainable buildings. *In:* 2010 4th Annual IEEE Systems Conference, 5-8 April, San Diego, USA, pp. 277-282.

ZI-NING, Z., QING-SHAN, J., CHEN, S. & XIAOHONG, G. 2008. An Indoor Localization Algorithm for Lighting Control using RFID. *In:* IEEE Energy 2030 Conference, ENERGY 2008, 17-18 November, Atlanta, GA, USA, pp.1-6.

ZITTING, A. 1998. *Thermal Degradation Products of Polymers* [Online]. Stockholm, Sweden.: The Nordic Expert Group for Criteria. Available:

http://www.inchem.org/documents/kemi/kemi/ah1998_12.pdf [Accessed 04/04/2012].

APPENDIX A: Data handling

A.1 Launching HOBO data loggers

- Install HOBOware pro data logger software on a computer
- Connect data loggers to the computer via USB cable
- Set the chosen sampling time
- Launch loggers
- Connect loggers to indoor environmental sensors

A.2 Procedure for transferring data and storage

- Down load data from HOBO loggers connected to indoor environmental sensors
- Download weather data from Gateway building, and export data in to Microsoft Excel
- Export indoor environmental data in to Microsoft Excel
- Design adequate tables in MySQL data base
- Upload indoor environmental data in to MySQL database server
- Upload weather data in to MySQL database server
- Energy (electricity) data is automatically logged into MySQL database via a radio network

A.3 Procedure for gathering ground truth occupancy data

- Mount infrared camera on a desired position
- Install screen grabber software on laptop (Courtesy: www.theuds.com)
- Connect camera to laptop
- Adjust camera resolution until an optimal image is obtained
- Select sampling time for screen shot capture
- Download saved occupancy images
- Extract occupancy numbers from images in MATLAB environment
- Upload occupancy data in to MySQL database server

APPENDIX B: FLIR A40 camera (data sheet)

The FLIR A40 infrared camera is a high resolution camera that offers real-time imaging solutions for machine vision and remote monitoring applications. It features built-in logic and plug-and-play network configurations. Figure (B.1) illustrates the camera interfaces.

Features

- Precise thermal measurement
- Real-time digital video output
- Firewire or Ethernet connection options
- Maintenance-free, uncooled, microbolometer detector
- Multiple users can access data from multiple cameras
- LabView and C++ / Visual Basic support

Figure (B.1): Installed FLIR camera applied in the study

Figure (B.2): FLIR A40 camera interfaces (Courtesy: FLIR A40 manual)

Figure (B.2): FLIR A40 camera interfaces (Courtesy: FLIR A40 manual)

Table B.1: FLIR A40 infrared camera characteristics

System Type	Focal plane array
Spectral range	Long wave
Detector	320 X 240
Detector material	Microbolometer
Measurement accuracy	$\pm 2^{0}C$
Measurement range	$0\text{-}1500^{0}C$
With filter	2000C
Field view	$45^{0}C$
Cooling	Uncooled
Spatial	Lens dependent
Thermal sensitivity	<0.10 at $30\ ^{0}C$
Detector refresh rate	30Hz
Dynamic range	14bit
Emissivity adjustment	0.01-1.00
Palettes	Multiple
Display type	Video out to LCD monitor
Image storage capacity	100 images
Storage medium	On-board RAM
Operating temperature	$-20 – 50^{0}C$
Camera weight	< 1.5Kg
Focus distance	< 1 inch
Video output	Standard NTSC
Power supply	A/C power
Voice annotation	N/A
Available accessories	Batteries , tripods , software

Courtesy: FLIR A40 manual

APPENDIX C: Dual monostable precision IC (CD4583) data sheet

The CD4538BC is a 16-pin dual, precision monostable multi-vibrator that is free from false triggering. This device possesses an independent trigger and reset controls that are internally latched. This IC is wired as a monostable timer, with accuracy and pulse duration dependent on the timing components, R_t and C_t. The CD4538BC IC does not allow the timing capacitor, C_t, to discharge through the timing pin on power-down condition. Hence, no external resistor is required in series with the timing pin for protection. The output is high at power ON condition, and would transit to LOW when the trigger receives a low-to-high pulse. This IC offers a low cost alternative for designing timing circuits. Further information can be seen on CD4538 technical data sheet (Courtesy: www.fairchildsemi.com)

- **CD4538BC characteristics**

Separate latched reset inputs	
Wide supply voltage range	3.0V to 15V
High noise immunity	0.45Vcc (typ.)
Pulse-width variation from part to part	± 1.0%
Low standby current	5nA (tpy.) @ 5Vdc
Logical function	Monostable multivibrator
Technology	CMOS
Operating temperature range	-55^{0}C – 120^{0}C

APPENDIX D: MATLAB codes for data handling

D.1 Data upload to MySQL database

```
mysql('open','146.227.24.XX','toby_e','xxxxx')
mysql('use toby_e_test_area_3')

%fid = fopen ('c:\toby_phd_data_collection.csv')

working_dir = ['G:\test_area_3\']
filesindir = dir ([working_dir,'*.csv'])
    for i = 1:length(filesindir)
        filename = [working_dir,char(filesindir(i).name)]
        logger_file = char(filesindir(i).name)
        fid = fopen (filename)

    %file loop will go here

    instr = fgetl(fid);
    instr = fgetl(fid);

    %leaving out 2 header files
    looping = 1;

    while looping ==1

    instr = fgetl(fid)


    % 1, 07/12/12 02:11:01 PM,23.388,

        if length(instr)<3
%file ends for hoboware pro text files
            looping = 0;
            break
        end

    %now to pull out date and sensor data content from each line for
the text file
    commas = strfind(instr,(','));
        if length(commas) <3
%pad with comma for hoboware pro text files
            instr = [instr,','];
            commas = strfind(instr,(','));

        end

    %instr = fgetl(fid)
    sample_no = instr(1:(commas(1))-1);
    date_time = instr(commas(1)+1:commas(2)-1)
    CO2= instr(commas(2)+1:commas(3)-1);
    sound = instr(commas(3)+1:commas(4)-1);
```

```
    case_temp= instr(commas(4)+1:commas(5)-1);
    spaces = strfind(date_time,' ');
    slashes = strfind(date_time,('/'))
    colons = strfind(date_time,':');
    date = date_time(1:(slashes(1)-1));
    spaces = strfind(date_time,' ');
    month = date_time((slashes(1)+1:slashes(2)-1));
    year = date_time((slashes(2)+1):spaces(1)-1);
    hour = date_time(spaces(1)+1:colons(1)-1);
    minute = date_time(spaces(1)+4:spaces(1)+5);


    qct = ['''',',',''''];qt = [''''];

    mysqldate = [year,'-',month,'-',date,' ',hour,':',minute];
    exceldate = [date,'/',month,'/',year,' ',hour,':',minute];


    if instr(length(instr)-5:length(instr)) == 'Logged'   %file end
for Hoboware pre v2.2
        looping = 0;

    end
        %assemble the SQL

    matlabdate =
datenum(str2num(year),str2num(month),str2num(date),str2num(hour),str
2num(minute),0)

    matlab_date = num2str(matlabdate)


    insert1 = ['INSERT INTO
ltrh_4(sample_no,mysqldate,exceldate,matlab_date'];
    insert2 = [',CO2,sound,case_temp) '];
    values = ['VALUES
(',qt,sample_no,qct,mysqldate,qct,exceldate,qct,matlab_date,qct,CO2,
qct,sound,qct,case_temp,qt,');']



    sql_command = [insert1,insert2,values];
  a= mysql(sql_command);


    %disp([filename,' ',logger,' ',mysqldate,' ',CO2,’ ‘,sound,’
‘,case_temp,’])
    end

    fclose(fid)

end

mysql('close')
```

D.2 Data retrieval from MySQL database

```
mysql('open','146.227.XX,'toby_e','xxxxx')
mysql('use toby_e_ test_area_3')

%pull out co2 data
a= mysql('select co2 from CO2 where mysqldate between "2012-12-18
00:00:00" and "2012-12-19 00:00:00"');

%pull out sound data
b= mysql('select snd from sound where mysqldate between "2012-12-18
00:00:00" and "2012-12-19 00:00:00"');

%pull out case_temp data
d= mysql('select cas from case_temp where mysqldate between "2012-
12-18 00:00:00" and "2012-12-19 00:00:00"');

%Pull time data
e= mysql('select matlab_date from CO2 where mysqldate between "2012-
12-18 00:00:00" and "2012-12-19 00:00:00"');
```

D.3 Extract occupancy numbers

```
filepath = 'F:\Images_test_area_3'
filesout = 'F:\Images_test_area_3'

filename = 'F:\Images_test_area_3.csv'

fullfile = [filepath,'\',filename];

dirvar = [filepath,'\*.jpg']

filesindir = dir(dirvar)

%grab names of all files in directory (structure array)

sfiles = size(filesindir);    %how many fines in dir
sfilesloop = sfiles(1)

datafile = 'F:\Images.mat'

fid = fopen(filename,'w')

%times_array = cell(1,sfilesloop);

mysqldate = cell(1);
exceldate = cell(1);
exceltime = cell(1);
```

```
for imagecount = 1:sfilesloop

  %loop through and grab data for each image

    filename = char(filesindir(imagecount).name)    %index into
struct aray, get cell array, convert to char array

    % e.g. 18-12-2012_15-28-47.jpg
    %      12345678901234567890

    hour = filename(12:13);minute = filename(15:16);sec =
filename(18:19);
    year = filename(7:10); month = filename(4:5);   date =
filename(1:2);

 matlabdate(imagecount) =
datenum(str2num(year),str2num(month),str2num(date),str2num(hour),str
2num(minute),str2num(sec));
    mysqldate{imagecount} = [year,'-',month,'-',date,'
',hour,':',minute,':',sec];
    exceldate{imagecount} = [date,'/',month,'/',year];
    exceltime{imagecount} = [hour,':',minute,':',sec];

    A = imread([filepath,'\',filename],'jpg');
    image(A)
    countstr = num2str(imagecount)
    localcount = input(['Image ',countstr,'
',hour,':',minute,':',sec,' How many people? ']);

    if isempty(localcount) ==1
        people_count(imagecount) = 0;
    else
        people_count(imagecount) = localcount;
    end
    disp(people_count(imagecount))

        fprintf(fid,'%s, %s, %f\n',mysqldate{imagecount},
matlabdate(imagecount), people_count(imagecount));


end

mysqldate = mysqldate';
exceldate = exceldate';
exceltime = exceltime';
matlabdate = matlabdate';
people_count= people_count';

fclose(fid);
 %save G:\toby_phd\images_test_area_3\data matlabdate mysqldate
exceldate exceltime people_count

save (datafile2,
'matlabdate','mysqldate','exceldate','exceltime','people_count')
```

APPENDIX E: Occupancy for various sensor configurations on the 14th, 17th and 20th

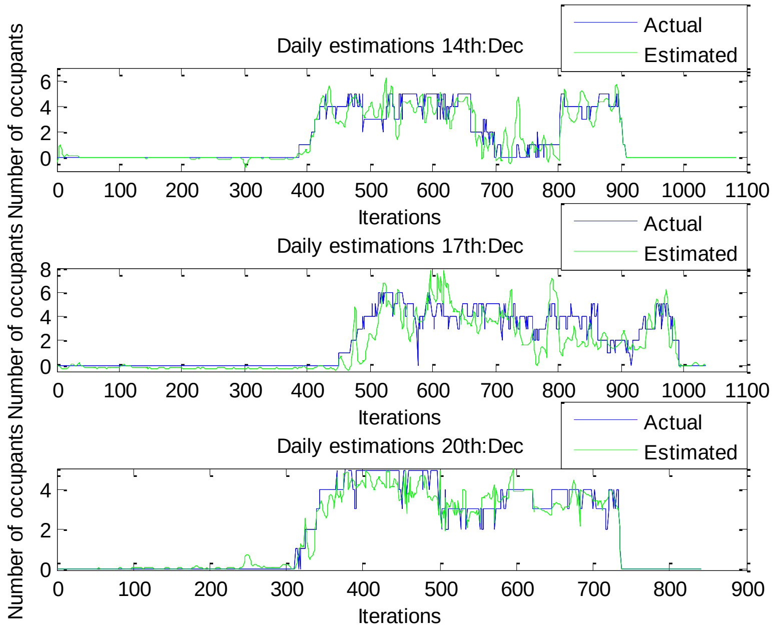


Figure (E.1): Occupancy results using heterogeneous multi-sensor network

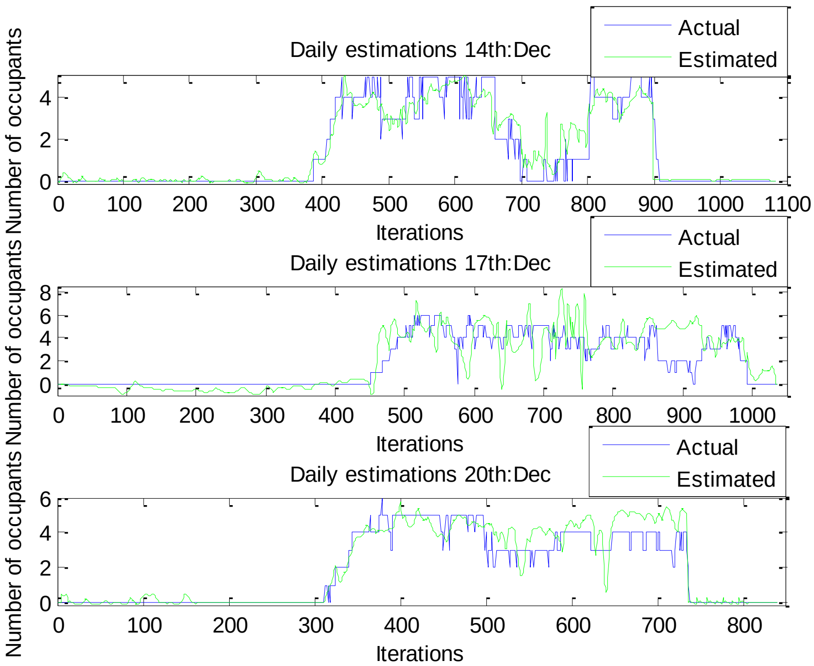


Figure (E.2): Occupancy results using CO_2 sensor network

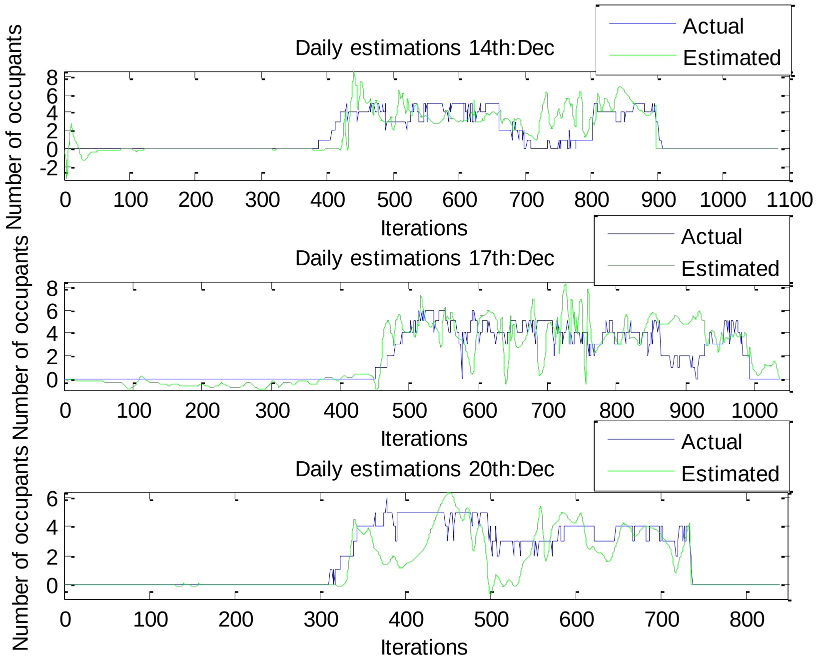


Figure (E.3): Occupancy results using case temperature sensor network

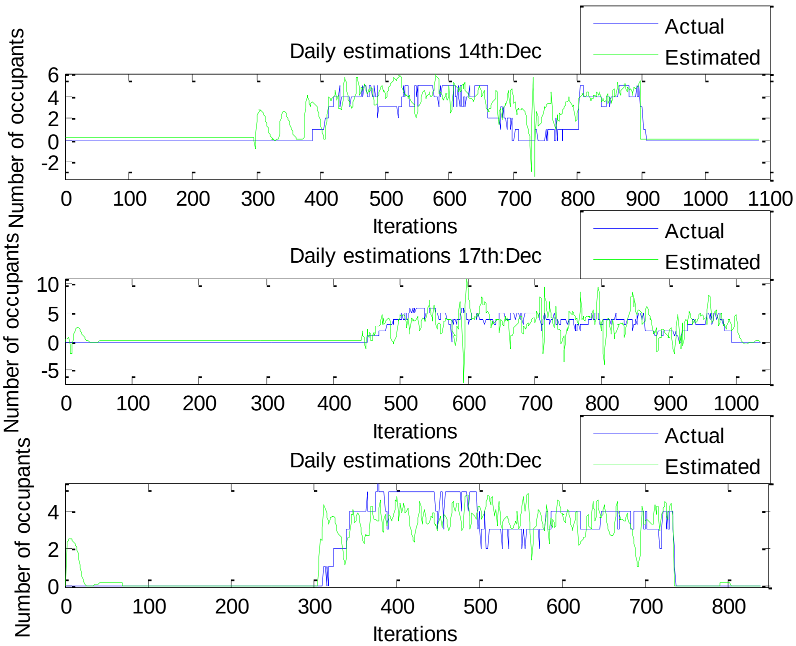


Figure (E.4): Occupancy results using PIR sensor network

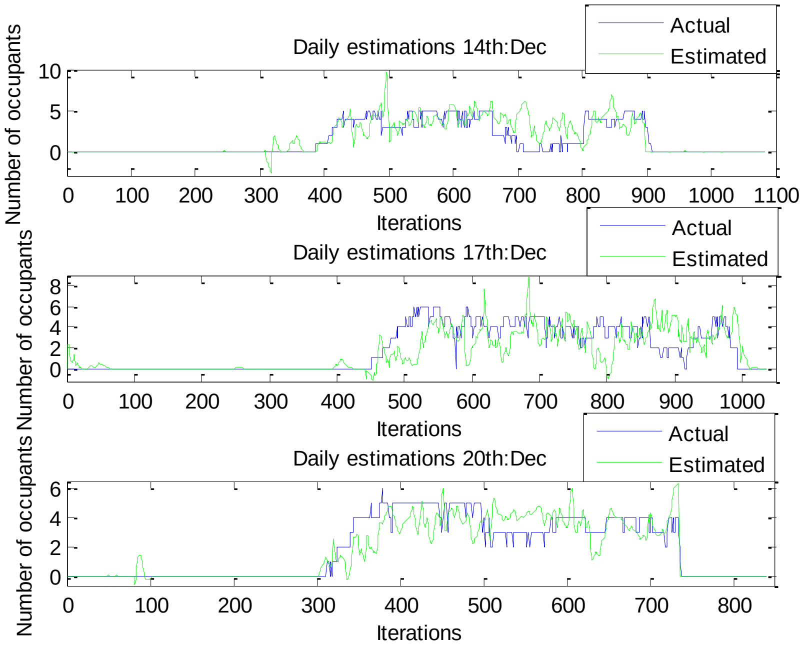


Figure (E.5): Occupancy results using sound sensor network

APPENDIX F: Comparison of models

Table (F.1): Comparison of models (14th)

Performance Metrics	NN	LR	SVM	RBF
RMSE	0.842	0.898	0.844	0.914
NRMSE	0.168	0.180	0.169	0.182
R^2	0.849	0.794	0.846	0.787
RAE	0.244	0.294	0.256	0.301
Accuracy	73.20	69.71	72.75	69.18

Table (F.2): Comparison of models (17th)

Performance Metrics	NN	LR	SVM	RBF
RMSE	1.161	1.180	1.098	1.189
NRMSE	0.194	0.197	0.183	0.213
R^2	0.707	0.699	0.750	0.690
RAE	0.349	0.356	0.302	0.358
Accuracy	62.24	61.16	68.49	60.44

Table (F.3): Comparison of models (18th)

Performance Metrics	NN	LR	SVM	RBF
RMSE	0.815	0.941	0.846	0.869
NRMSE	0.136	0.157	0.141	0.148
R^2	0.859	0.747	0.847	0.834
RAE	0.229	0.325	0.233	0.269
Accuracy	74.91	66.84	72.87	70.35

Table (F.4): Comparison of models (19th)

Performance Metrics	NN	LR	SVM	RBF
RMSE	1.064	1.046	0.925	1.016
NRMSE	0.152	0.149	0.132	0.145
R^2	0.827	0.834	0.840	0.839
RAE	0.288	0.298	0.240	0.242
Accuracy	68.53	69.88	71.95	70.72

Table (F.5): Comparison of models (20^{th})

Performance Metrics	NN	LR	SVM	RBF
RMSE	0.845	0.913	0.851	0.860
NRMSE	0.141	0.152	0.142	0.143
R^2	0.830	0.802	0.826	0.823
RAE	0.246	0.297	0.255	0.257
Accuracy	71.89	68.70	70.64	70.43

APPENDIX G: Linear regression model for the typical week tested

Linear Regression Model-14th

$\hat{O} =$ 0.2908 * AVR_CAS + 0.1338 * VAR_CAS + 3.4708 * FDIFF_CO_2 -2.3706 * AF_DIFF_CO_2 + 0.0043 * TOS_SND + 0.066 * VOS_SND -5.7646

Linear Regression Model- 17th

$\hat{O} =$ 0.2814 * AVR_CAS -0.2292 * VAR_CAS + 3.4479 * FDIFF_CO_2 -2.5669 * AF_DIFF_CO_2 + 0.0064 * TOS_SND + 0.0762 * VOS_SND -5.3694

Linear regression model-18th

$\hat{O} =$ 0.2932 * AVR_CAS -0.2506 * VAR_CAS + 3.2807 * FDIFF_CO_2 -2.3487 * AF_DIFF_CO_2 + 0.0069 * TOS_SND + 0.068 * VOS_SND - 5.5864

Linear Regression Model- 19th

$\hat{O} =$ 0.2402 * AVR_CAS -0.2159 * VAR_CAS + 3.7838 * FDIFF_CO_2 – 2.4483* AF_DIFF_CO_2 + 0.0069 * TOS_SND + 0.0853 * VOS_SND -4.6558

Linear Regression Model- 20th

$\hat{O} =$ 0.2814 * AVR_CAS - 0.2311 * VAR_CAS + 3.4782 * FDIFF_CO_2 -2.5871 * AF_DIFF_CO_2 + 0.0066 * TOS_SND + 0.0781 * VOS_SND -5.472

APPENDIX H: Complete set of sensory inputs and optimal sensory inputs

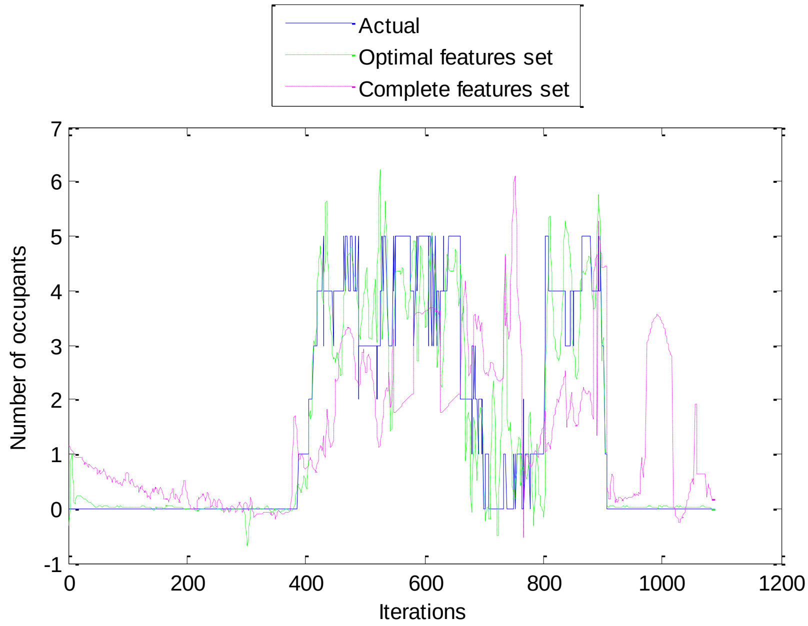


Figure (H.1): Complete and optimal multi-sensory features- 14th

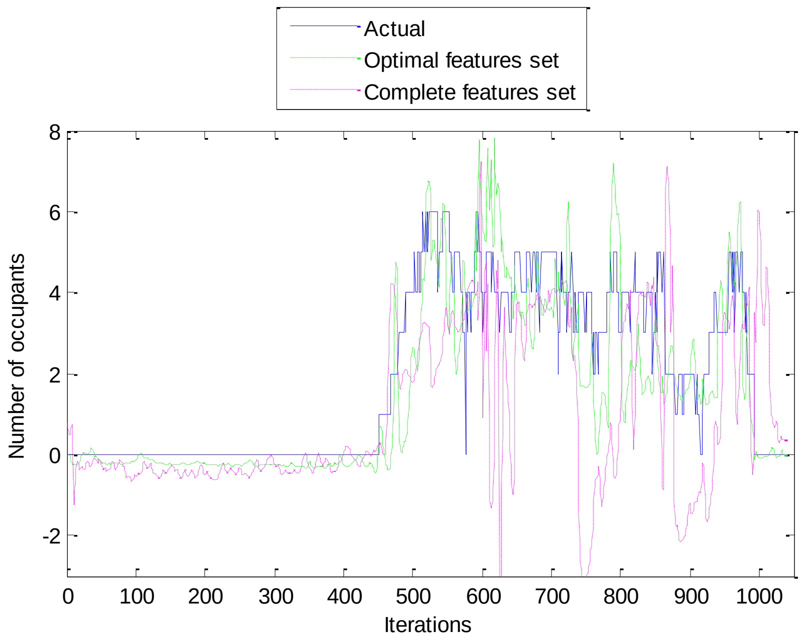


Figure (H.2): Complete and optimal multi-sensory features- 17th

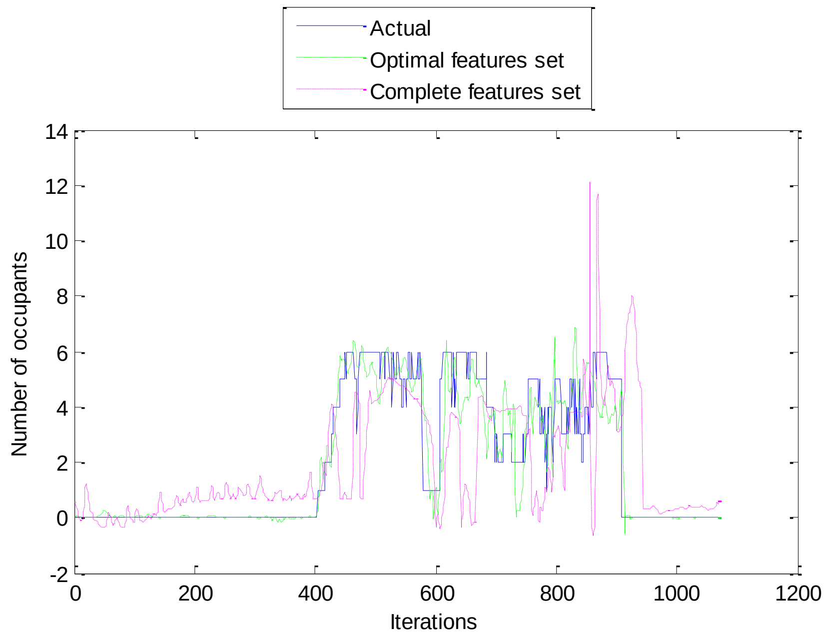


Figure (H.3): Complete and optimal multi-sensory features- 18th

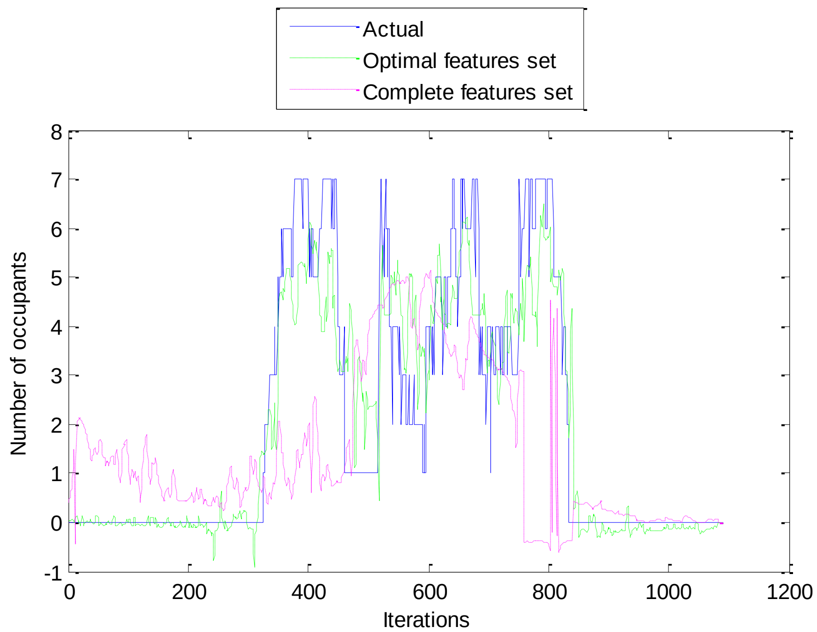


Figure (H.4): Complete and optimal multi-sensory features- 19th

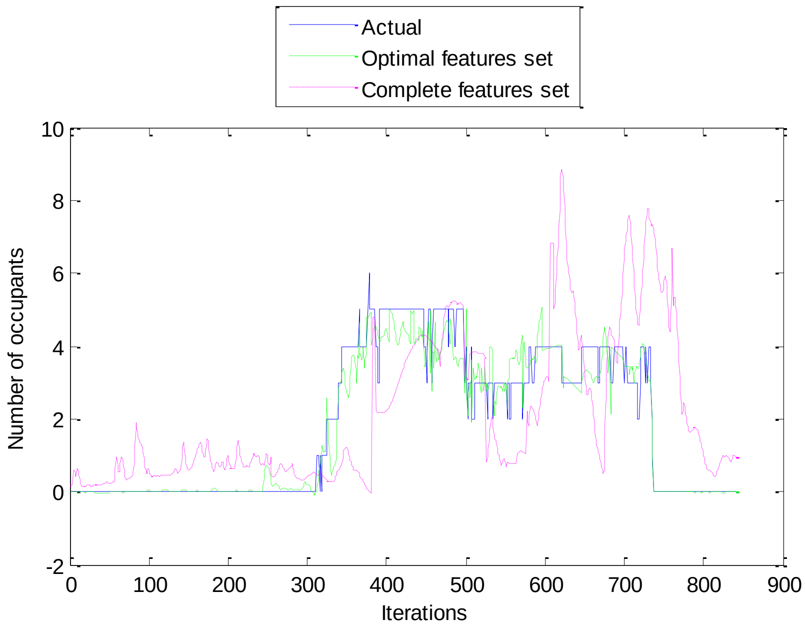


Figure (H.5): Complete and optimal multi-sensory features- 20th